

THE
BRITISH JOURNAL
OF
STATISTICAL
PSYCHOLOGY

EDITED BY
CYRIL BURT

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND J. W. WHITFIELD

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN	L. S. PENROSE
M. S. BARTLETT	J. FRASER ROBERTS
W. G. EMMETT	A. RODGER
M. G. KENDALL	W. STEPHENSON
D. N. LAWLEY	A. STUART
E. S. PEARSON	P. E. VERNON

Managing Editor : H. G. MAULE

Volume VII
1954

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4.





Digitized by the Internet Archive
in 2026

A FACTORIAL ANALYSIS OF ATHLETIC ABILITY

By G. HIGHMORE and W. R. TAYLOR

Loughborough Training College¹

I. *Data.* II. *Factorial Analyses.* III. *Estimation of General Athletic Ability.* IV. *Summary and Conclusions.*

I. DATA

Problem. The object of the enquiry described in the following paper was to attempt an analysis of athletic ability among schoolboys and to devise suitable tests for that purpose. The problem is of practical as well as theoretical interest. Although records of athletic prowess are often passed on to the secondary school from the primary when pupils are transferred, there is still a need for some simple procedure which will enable teachers of physical education to obtain a quick preliminary estimate of the characteristics of individual children who for the first time come under their observation.

The earliest research which applied correlational techniques to various test-results—Wissler's well-known investigation on general and special abilities published in 1901—included tests of physical as well as mental ability, and was based on the scheme drawn up by Galton and J. M. Cattell (1 and refs.). One of the main conclusions was that "the physical tests show a general tendency to correlate amongst themselves, but only to a very slight degree with the mental tests." The 'general tendency' thus observed provides some measure of presumptive evidence for what may be called a 'general physical ability'; but, although the complete scheme apparently included what were termed 'gymnasium tests,' no systematic endeavour seems to have been made to measure athletic abilities as such. Some years later Reilly (3) made out a strong case for the classification of school pupils according to their athletic capacities, and outlined a scheme on a 'rational' or *a priori* basis.

The first attempt to apply factorial methods to such measurements seems to have been the enquiry carried out by Burt and his co-workers in several London schools, and reported in one of his L.C.C. Memoranda (4; cf. 11, p. 111): he found a well-marked general factor, closely correlated with general physical health, and special factors for strength, quickness, dexterity, steadiness, and (later) endurance. Subsequently, at the Institute of Education both he and Professor Hamley initiated several minor investigations which were carried out in London schools by research-students under their joint supervision. In the United States Cozens was apparently the first to apply correlational techniques to direct measurements of athletic capacity

¹ Special acknowledgements are due to the various members of the Physical Education Research Group of the College, and especially to Mr. H. T. A. Whiting (their Secretary) and Mr. N. France (Senior Lecturer in Education and Mathematics), all of whom gave great assistance with the measurements and calculations, under the general supervision of Dr. Highmore; also to Mr. C. H. Moore, Headmaster of the Limehurst School, Loughborough, where the investigations were carried out, and to Mr. S. G. Jarvis, the master in charge of Physical Education. The paper in its present form owes much to Mr. W. R. Taylor, who, though not a member of the College, has worked over the original factorial analysis and helped in preparing the original draft for publication in this *Journal*: his name has therefore been added as joint author. Further details, tables, and charts will be found in the cyclostyled report by the Research Group, obtainable from the College (13).

(see 6 and later papers by the same author for excellent discussions of American work). Of more recent investigators in that country the numerous studies by McCloy deserve special mention (7, 9). He finds three main common factors—strength, speed, and dead weight. However, he was apparently already convinced of the existence of these factors, and the analysis with its rotations seems largely called in to provide a convenient confirmation of what he believed on more general grounds (for a detailed account with criticisms, see 12, p. 269). More recently, on the initiative of Professor Hamley, Dr. Highmore has carried out research on the subject, with the assistance of Mrs. Edwards-Penfold (12); while a somewhat similar study of the physical capacities of male adults has been described by Vernon in an unpublished memorandum (10).

Subjects and Tests. The subjects of the present investigation were 110 eleven-year-old boys attending a secondary modern school. After some preliminary experiments, the following tests were selected and standardized. A full description of the procedures, with means, standard deviations, and frequency-distributions, will be found in the memorandum already cited (13).

1. A 50-yard sprint on the grass playing field. 2. A run of 150 yards made on a gently curving track. 3. A standing upward jump. 4. A standing long jump; in the case of tests 3 and 4 three jumps were allowed, and the best was recorded. 5. A high jump: three trials were allowed at each height, and a success was recorded if the boy succeeded in one trial out of three. 6. Throwing the cricket ball: a 5½-oz. ball was used, and a preparatory run allowed. 7. Football throw-in: three throws were made, and the best recorded. 8. Football punt for distance: a dead association football (No. 4 size) was kicked from a marked line, and the best of three kicks accepted.

Before the tests were carried out, the senior physical education master and his assistant were asked to place the boys in order of merit according to the all-round performance of each in the traditional track and field events. This was provisionally accepted as the criterion. In addition, each boy was measured for height and weight.

II. FACTORIAL ANALYSES

Inter-Correlations.—The correlations of the tests with each other and with the criterion are set out in Table I. With a group of this size, any coefficient of .18 and over will be more than twice its standard error. Thus all the correlations in the table are statistically significant. In a further investigation it might be better to factorize the results of the tests by themselves, and then compute the correlations between the factors and the criterion furnished by the teachers' independent assessments. For purposes of a preliminary study, however, it is simpler to start by including the criterion in the table to be factorized.

TABLE I. CORRELATIONS BETWEEN TESTS

Test	1	2	3	4	5	6	7	8	9
1. Sprinting (50 yards)	(.63)	.56	.51	.43	.51	.42	.36	.30	.50
2. Long Jump	.56	(.58)	.44	.41	.41	.36	.41	.34	.38
3. Running (150 yards)	.51	.44	(.43)	.39	.40	.21	.23	.18	.40
4. High Jump	.43	.41	.39	(.47)	.37	.40	.40	.37	.46
5. Standing Upward Jump	.51	.41	.40	.37	(.40)	.21	.26	.22	.34
6. Throwing Cricket Ball	.42	.36	.21	.40	.21	(.57)	.49	.53	.52
7. Punting Football	.36	.41	.23	.40	.26	.49	(.45)	.41	.52
8. Throwing Football	.30	.34	.18	.37	.22	.53	.41	(.58)	.57
9. Criterion	.50	.38	.40	.46	.34	.52	.52	.57	(.63)

Three methods of analysis were adopted.

1. *Simple Summation.* A preliminary analysis was made by the ordinary method of simple summation. The self-correlations (or 'communalities') were obtained by successive approximation: three iterations proved sufficient, and the resulting estimates are shown in brackets in Table I. Three factors were extracted. The saturations are shown in Table IIA. Instead of following the usual rule of arranging the tests in order of minimal sign-changes, we have adopted an arrangement which indicates more clearly the separate subdivision of the first bipolar by the second.

The first is a large general factor accounting for nearly 42 per cent. of the variance, far more than the remaining factors put together. This result agrees with that reported by Burt, who also found "a large general factor accounting for nearly half the total variance." It fully justifies the attempt made, when planning the enquiry, to include an all-round assessment of athletic ability as a criterion; and, since the correlation (or 'saturation') of that criterion with the general factor is higher than that of any of the eight tests, we may regard the two as approximately identical. The tests having the highest correlation with this factor are (in order) the 50-yard sprint, the long jump, throwing the cricket ball, the high jump, and kicking and throwing the football: all have saturations over 0.6.

The second factor is bipolar, and accounts for 8.5 per cent. of the variance. It appears to distinguish what may be called the sport features from the games features; but the suggestion must be made with some reserve. The third factor accounts for 1.5 per cent. of the variance, and is only just significant. It suggests a subdivision of the various activities into three main groups, or possibly four if we include the criterion.

2. *Non-Overlapping Group Factors.* With only 110 persons in the sample, it would scarcely seem legitimate to attempt the extraction of as many as four group factors in addition to the general or basic factor. This was confirmed by actual trial. Accordingly, a preliminary analysis into three non-overlapping group factors was next undertaken (the formula used was that given by Burt in *Factors of the Mind*). The lines of division were those indicated by the preliminary summational or 'bipolar' analysis. The resulting saturations are shown in Table IIB.

3. *Analysis of Overlapping Group Factors.* The residuals from the last analysis indicated a good deal of overlapping. Accordingly, with the matrix of non-overlapping group factors as a starting point, a set of overlapping group factors (including one general) was obtained by means of an arithmetical rotation: the method adopted was that described in this *Journal* (III, p. 75). All except two of the negative saturations were thus suppressed. By means of successive approximation these were reduced to negligible size, and a slightly better fit was at the same time achieved.¹ The saturations thus obtained are shown in Table IIC: those under 0.200 are statistically

¹ A graphical rotation was first attempted, starting with the 'bipolar' (or 'centroid') matrix of four factors. The 'bipolar' factor-matrix can evidently be regarded as a sum-and-difference matrix obtained from the positive group factor-matrix. Hence, to reverse this process, a preliminary arithmetical rotation was made by post-multiplying the summational factor-matrix by a 'transclassificatory operator'—in this case a 3×4 semiorthogonal matrix of the 'equilibrated' type (cf. this *Journal*, I, p. 8 and refs.). In this case the rotation matrix adopted was the following:

·904	·247	·247	·247
·000	·707	—·707	·000
·000	·408	·408	—·816

The first figure in the first row is the ratio

$$\frac{\text{average saturation for basic factor}}{\text{average saturation for general factor}} = \frac{a}{g}$$

and the remaining figures are inserted to bring the sum of the squares to unity. If a group factor analysis has not been employed, we can take $a = \sqrt{(g^2 - 2h^2)}$ and $b = 1.414h$, where g denotes the

non-significant. A reconstruction of the original correlations showed that an excellent fit was thus obtained with one basic and three group factors only. The majority of the discrepancies were numerically less than 0.030, and all were well below the standard error.

An attempt was also made to secure a factor-matrix showing 'simple structure' by means of the rotation-method described by Thurstone. This was only possible when 'oblique' or highly correlated factors were introduced; and that in turn suggests the desirability of computing 'second order' factors, based on the saturations for 'simple structure.' The final result resembled that shown in Table IIC without the general factor; but the figures furnished a comparatively poor fit. In itself a 'simple structure' would seem somewhat unsatisfactory as a final solution. To begin with, several 'simple structures' prove to be possible, and the choice depends on the arbitrary choice of the computer. With all of these the 'basic' factor, representing general athletic ability (in which we are primarily interested), necessarily disappears, and the group factors show little relation to the classification indicated by the bipolar matrix. Thus the ultimate effect is to suppress a factor which accounts for far more of the variance than all the rest put together, and to re-interpret the resulting correlations in terms of a variable set of correlated group factors.

The analysis into a basic factor supplemented by three group factors is in keeping with the results obtained by Burt (who found five factors), by Vernon, and by Highmore respectively, in earlier enquiries. The proportions contributed to the total variance are shown in Table III below.

TABLE III. RELATIVE SIZE OF BASIC AND GROUP FACTORS
(IN PERCENTAGES)

Investigation	Basic Factor	First Three Group Factors
Present Data	30.2	23.7
Burt	35.4	18.8
Vernon	27.4	16.4
Highmore	32.0	11.7

average general factor-saturation, and h the average numerical value for the saturations of the larger bipolar factor. The values for the first row can then be computed from the expressions

$$a/\sqrt{(a^2 + \frac{1}{3}b^2)}, \quad b/\sqrt{(a^2 + \frac{1}{3}b^2)},$$

where a denotes the average basic factor-saturation and b the average group factor-saturation (which are provisionally treated as equal). In either case the figures in the other rows of the rotation matrix are constants, viz., $\sqrt{(1/2)}, \sqrt{(1/6)}, \sqrt{(2/3)}$. The rotated factor-matrix so obtained was then subjected to further graphical rotations, taking one pair of factors at a time; and these were continued until all negative saturations were reduced to negligible size. The final results were very similar to those given in Table IIC. In actual practice this procedure is both quick and easy; but, in a research like the present, owing to the crudeness of all merely graphical procedures, it is not entirely satisfactory.

Accordingly, the figures finally given were obtained by the slower but more accurate method of successive approximation. The usual working procedure was adopted. The method of carrying out the approximations was that which is now in general use. Taking the results of the arithmetical rotation as a fresh starting point, a correlation matrix was constructed from all except one of the factors, namely, the first: this reconstructed matrix was then subtracted from the matrix of observed correlations; and improved values for the first factor were then computed from the matrix differences by the ordinary method of simple summation. A new correlation matrix was next constructed, this time missing out the second factor; and improved saturations for this factor were obtained in the same manner. In this way all four factors were adjusted in turn. A final check was made by applying Burt's 'tangent formula' for minimizing the squares of non-significant saturations (virtually the same as maximizing fourth moments: *Factors of the Mind*, pp. 353-4). The changes thus indicated proved negligible. On making a final reconstruction of the correlation table, it was found that these figures gave a much closer fit to Table I than those obtained by the graphical procedure.

TABLE II. FACTOR-SATURATIONS OBTAINED BY VARIOUS METHODS

Type of Factor Method of Computation		A. Bipolar Factors Simple Summation		B. Non-Overlapping Factors Direct Group Factor Analysis		C. Overlapping Group Factors Arithmetical Rotation	
1. (D) Sprinting (50 yards)	..	..	..	..	..	..	..
4. (J) Long Jump ..	..	..	..	..	..	..	..
Subtotal	..	..	..	..	..	..	..
2. (E) Running (150 yards)	..	..	..	..	..	..	..
5. (K) High Jump ..	..	..	..	..	..	..	..
3. (I) Standing Upward Jump ..	..	..	..	..	..	..	..
Subtotal	..	..	..	..	..	..	..
6. (G) Throwing Cricket Ball	..	..	..	..	..	..	..
8. (H) Punting Football ..	..	..	..	..	..	..	..
Subtotal	..	..	..	..	..	..	..
7. (F) Throwing Football	..	..	..	..	..	..	..
9. (L) Criterion ..	..	..	..	..	..	..	..
Subtotal	..	..	..	..	..	..	..
Sum of squares	..	..	..	..	..	..	..
Contribution to variance (per cent.)	..	..	..	..	..	..	..

The low value for the basic factor found in Vernon's study of adults may at first sight appear to be consistent with the view that physical capacities, like mental, tend to become more differentiated and specialized with increasing age; but it seems more probable that the chief cause of the differences is the difference in the heterogeneity of the several groups.

Interpretation of Final Group Factor-Matrix. The first factor accounts for 30 per cent. of the variance. It is evidently a general or basic factor underlying athletic performances generally. The criterion has the highest saturation for this factor.

Of the broad group factors, the first and largest appears to class together the five tests of sporting events: those showing the highest saturations are the 50-yard sprint and the 150-yard run, while the high jump shows the lowest (barely significant). The second group factor classes together the four games tests: here the two throwing tests have the highest saturations. The last group factor is much smaller, and suggests a subdivision of the sport tests somewhat like that indicated by the last bipolar factor: the two tests having the largest saturations are the high jump and the standing upward jump, though the saturation for running is almost as large. The criterion has a moderately large saturation for both the last two factors.

The resulting group factors show some resemblance to those found in the earlier L.C.C. investigations. In addition to the usual type of test based on field events, these also included laboratory tests of simpler physical capacities, taken from Whipple (2), and assessments for muscular development and general physical condition based on assessments supplied by the school medical officer. As already noted, a well-marked general or basic factor was found in every case, which appeared to be highly correlated with a general factor for health: and it is further stated that "both in tests of athletic performance and in tests of a laboratory type we have found distinguishable factors for (i) strength, (ii) quickness, (iii) skill or dexterity, and (iv) steadiness" (11, p. 111), to which later on a factor of (v) "endurance" was added (4, p. 18). The tests of short-distance running were highly saturated with a factor identified as 'motor speed'; the tests of long-distance running involved in addition a factor of 'endurance'; putting the weight, and to a less extent throwing and jumping, involved a factor of 'muscular strength'; throwing appeared also to involve a factor of 'muscular dexterity' (knack as well as skill in aiming). Among older youths the contribution of the general factor appeared to diminish, that of the group factors to increase; while one or two further factors appeared to emerge—e.g., one contrasting strength of leg and strength of arm, and another contrasting cricketers with footballers—the last (the footballers' factor) being positively correlated with relative weight. In general, height was correlated positively with running; relative weight (i.e., deviation of actual weight from expected weight for age and height) was correlated negatively with jumping, but the correlations were only slight.

In the present enquiry, it may be observed, the weight of the boys showed a small negative correlation with the standing upward jump, a moderately large positive correlation with football throw-in and football punt, and very small positive correlations with the rest. Height had a moderately large positive correlation with the long jump and the high jump.

American writers do not, as a rule, admit a general factor; but the more specialized abilities which they describe resemble those found by the British investigators. To assess athletic ability, however, they prefer in a large measure to rely on 'rational' or *a priori* considerations: thus MacCurdy (8) put forward a 'Physical Capacity Index,' which he calls 'Power,' and proposes to measure it by 'Force $\times$ Velocity': Force presumably is to be measured by jumping or putting the weight; Velocity by running. McCloy suggests a somewhat similar formula. He states that "in the analysis of many of the motor skills used in physical education and athletics, three orthogonal factors usually emerge: (1) strength, (2) speed, and (3) dead weight" (9).

III. ESTIMATION OF GENERAL ATHLETIC ABILITY

For practical purposes it would be desirable to obtain a single, all-round estimate for each individual, were such an estimate justifiable. Since here the first summational factor accounts for over 40 per cent. of the variance and both the basic factors account for over 30 per cent., this assumption appears to be fully warranted ; and, as we have seen, it is further borne out by the evidence of the provisional criterion. It is moreover noteworthy that the criterion itself also shows positive saturations for the three supplementary group factors, though only those for the last two factors appear to be fully significant. We should therefore perhaps regard it as in some ways a composite assessment and possibly a compromise. There seems little doubt that assessments based on objective tests would be at once more reliable and more valid than impressionistic estimates like those which have formed the provisional criterion.

Since the eight tests employed fall into three fairly distinct groups, it may be inferred that moderately good multiple correlations would be obtained by selecting one test only from each of the groups, i.e., by using at most three tests instead of eight. The first test to be chosen will obviously be the test having the highest correlation with the first summational factor, namely, the 50-yard sprint. This test, however, is rather over-weighted with the first of the group factors. Hence it would be undesirable to include the long jump as well, in spite of its high saturation, since it has much the same factorial composition. The next to take must therefore be the football throw-in : this not only has the highest saturation for the other large group factor, but also (as may be seen from the bipolar analysis) will act to some extent as a suppressor variable, counteracting the bias of the sprinting test. The multiple correlation with the criterion furnished by these two tests amounts to 0.64. If for the third test we choose the high jump, which has a high general saturation with the basic factor and a higher saturation for the third group factor than any of the other tests, the correlation is raised to 0.68. Adding a fourth test only raises it to 0.69 ; and the improvement is hardly enough to justify the additional labour a fourth test would entail.

The regression equations show that, if two tests alone are employed, the sprinting test should be given six or seven times as much weight as the football throw-in test. If, however, three tests are employed, the regression coefficients differ but slightly ; and an excellent approximation is obtained by giving each test equal weight, e.g., by ranking the pupils in order of merit for every one in turn, and then averaging the orders, or (if only a few individuals are tested at a time) reducing the raw measurements to standard measure and then averaging the reduced figures.

On the whole, therefore, it would seem that a tolerably good estimate of all-round athletic ability can be obtained with two or three tests only. Those whose interest lies chiefly in practical applications will find detailed suggestions regarding administration, scoring, and classification in the memorandum already quoted (13).

IV. SUMMARY AND CONCLUSIONS

1. Eight tests of athletic ability have been devised for 11-year-old schoolboys, and applied to a sample consisting of 110 individuals. All the correlations were found to be significant and positive.

2. An analysis by simple summation reveals a large general factor accounting for nearly 42 per cent. of the variance and two bipolar factors subdividing the tests into three main groups.

3. The factorial composition of such tests seems to be more suitably expressed in terms of a set of basic and group factors. In addition to a factor for general athletic ability, three overlapping group factors were obtained which very tentatively might be described as factors for (i) sporting events (especially running), (ii) games tests (especially throwing), and (iii) jumping. Their relations to the factors reported by other investigators have been briefly discussed.

4. It is suggested that a good all-round estimate of a boy's athletic ability might be obtained by subjecting him to three selected tests—sprinting, football throw-in, and high jump, giving equal weight to each.

REFERENCES

1. Wissler, C. (1901). *The Correlation of Mental and Physical Tests* (Psych. Rev. Mon. Sup., III) New York : Macmillan.
2. Whipple, G. M. (1910). *Manual of Mental and Physical Tests*. Baltimore : Warwick and York.
3. Reilly, F. J. (1918). 'A Rational Classification of Boys and Girls for Athletic Competition.' *Am. Phys. Educ. Rev.*, XII, 58-71.
4. Burt, C. (1925). *Tests and Assessments of Physical Capacities*. London County Council.
5. Rogers, F. R. (1927). *Tests and Measurement Programmes in the Redirection of Physical Education*. New York : Teachers College, Columbia.
6. Cozens, F. W. (1929). *The Measurement of General Athletic Ability in College Men*. Oregon : University Press.
7. McCloy, C. H. (1932). *Measurement of Athletic Power*. New York : A. S. Barnes.
8. MacCurdy, H. L. (1933). *Test for Measuring the Physical Capacity of Secondary School Boys*. New York : the Author.
9. McCloy, C. H. (1939). *Tests and Measurements in Health and Physical Education*. New York : Crofts.
10. Vernon, P. E. (1945). *An Investigation of Tests of Physical Capacities*. Unpublished memorandum.
11. Burt, C. (1949). 'The Structure of the Mind : A Review of the Results of Factor Analysis.' *Brit. J. Educ. Psychol.*, XIX, 176-99.
12. Highmore, G. (1949). *A Factorial Analysis of Athletic Ability*. Unpublished Ph.D. Thesis. University of London.
13. Highmore, G. (1953). *An Analysis of Athletic Ability*. Unpublished memorandum obtainable from the author.

THE SIMPLEX STRUCTURE OF THE PROGRESSIVE MATRICES TEST

By K. R. GABRIEL

Statistical Laboratory, Hebrew University, Jerusalem

I. Description of the Test and the Experimental Group. II. Guttman's Simplex Theory. III. Analysis of the Correlation Matrix and Derived Matrices. IV. The Value of Simplex Analysis and its Use in Prediction.

I. DESCRIPTION OF THE TEST AND THE EXPERIMENTAL GROUP

The Progressive Matrices Test, as described by Raven (2), was administered to 830 male and female recruits to the Israel army.¹ It was given as a group paper-and-pencil test with a time limit of 20 minutes. The speed requirement was emphasized, but many of the testees did not reach the end of the test. Scoring was by number of correct items.

Raven's form of the test consists of five sets with 12 incomplete patterns in each, and the testee has to complete each pattern by choosing one of 6 or 8 alternatives. There is a rough order of difficulty within each set, but the sets differ somewhat in their contents. The successive sets are described as involving: (A) Continuous Patterns (1×1 matrices), (B) Analogies of Figures (2×2 matrices), (C) Progressive Development of Figures, (D) Arrangements and Combinations of Figures, and (E) Resolving Figures into their Constituent Parts (all 3×3 matrices). Though the evolution of the individual patterns cannot be described quite so simply, a general order of increasing complexity can certainly be observed among the sets. Indeed, the original intention in devising the test was, we are told,² to select a series of problems increasing step by step in complexity—complexity being defined at the outset on more or less logical grounds. The guiding principle was to present first a set of items involving no relation-finding at all (in order to familiarize the subject with the general notion of fitting insets), and then to start with a simple non-verbal form of analogies test—the 2×2 type of 'matrix' that had already been found to yield an effective test of intelligence (1, pp. 234–8, 2nd ed.), and so proceed to larger matrices, progressively increasing the number of relations introduced at each stage.

¹ The work described in this article was carried out under the auspices of the Israel Defence Army. My thanks are due to the staff of the Army's Psychotechnics Institute for providing the data and assisting me in the laborious computations required.

² The principles originally proposed for the construction of the Progressive Matrices test have been explained in some detail by Burt (4), who apparently suggested the method and expressed by algebraic formulæ the ideal scheme for systematically increasing complexity of the successive items.

II. GUTTMAN'S SIMPLEX THEORY

The structure of a battery of tests arranged in order of complexity has been studied by Professor Louis Guttman (7, 10). His Simplex structure postulates a series of elemental components of ability, each of which involves all the previous components in the series. Every test in the Simplex measures a section of this series; and its complexity depends on the extent of the series covered. The complexity of a test is thus indicated by the last member of the component series it measures, since all the preceding components must also have been included.¹ It follows that a test measures all the components of any test less complex than itself, plus a number of additional components, and that it measures only a portion of those components that will enter into the more complex tests.

A clear distinction must be made between complexity and difficulty of test items. Within each test, at any given level of complexity, items may still vary in difficulty.² Thus there may be difficult questions of a simpler nature, and easy questions of a more complex nature. We may, for instance, have difficult problems in addition and easy problems in multiplication.

The correlation matrix of a Simplex has certain properties which can be derived from the postulated formation of tests as cumulative functions of the elemental components.³ Evidently any particular test will correlate most highly with tests of a similar complexity, as these involve mostly the same components. The greater the difference in complexity, the smaller the extent of components shared by the tests, and therefore the lower the correlation. Thus in a Simplex correlation matrix in which the tests are arranged by order of complexity, the highest correlations in each row are found near the diagonal, and the correlations diminish the farther they are from the diagonal.

The properties of the multiple regression equations of each test on all the other tests, and of the inverse correlation matrix, which follow from these conditions, prove to be in many ways analogous. Every test can best be predicted by an equation including as many of its components as possible and as far as possible nothing else. Hence its regression equation will be confined to its neighbours taken in the order of complexity; the regression weights for more distant tests will be negligible. The inverse correlation matrix will consequently have high negative values near the diagonal, and zero or negligible values farther away from the diagonal.

¹ Burt has pointed out that, in theory, the original Binet-Simon Scale was constructed on an analogous principle, since success in any one of the later items is presumed in general to imply success in all the earlier items: cf. the schematic 'Answer Pattern' set out in the preceding issue of this *Journal*, Table I, 9, p. 7.

² The actual difficulty of the items in Raven's version has been empirically determined by Keir with the aid of the probit method (3, p. 146, Fig. 1). She finds that the increase in difficulty is by no means uniform, since there is a marked bunching of items of medium difficulty, and much overlapping of sets. Set D turns out to be on the whole slightly easier than set C; and within each set the numbering does not always represent the actual increase in difficulty. This conclusion has been confirmed by Banks and Sinha in a later issue of this *Journal* (6). They point out that the test could be much improved if the proposed principles for increasing complexity had been more rigorously adhered to, so as to secure a more even grading of the scale.

³ For the mathematical analysis of the Simplex and the proof of these and other properties, see Professor Guttman's article on Radex theory (10).

III. ANALYSIS OF THE CORRELATION MATRIX AND DERIVED MATRICES

Professor Guttman suggested that his Simplex theory might be used to test the hypothesis that the Progressive Matrices test exhibits a structure of increasingly complex sets. The inter-correlation matrix—Table I—was therefore computed, and it was found that the correlations sloped away from the diagonal, as would be expected in a Simplex.¹

TABLE I. CORRELATIONS BETWEEN SETS (N = 830)

Set	A	B	C	D	E
A	1·000	·567	·533	·472	·321
B	·567	1·000	·637	·539	·353
C	·533	·637	1·000	·656	·391
D	·472	·539	·656	1·000	·515
E	·321	·353	·391	·515	1·000

A more stringent test is provided by inspecting the inverse correlation matrix (Table II). This also showed that the larger elements were concentrated near the diagonal.

TABLE II. INVERSE MATRIX OF CORRELATIONS
BETWEEN SETS

Set	A	B	C	D	E
A	1·623	—·568	—·344	—·190	—·087
B	—·568	1·966	—·749	—·255	—·087
C	—·344	—·749	2·264	—·889	—·054
D	—·190	—·255	—·889	2·112	—·588
E	—·087	—·087	—·054	—·588	1·383

Some of the other elements, however, especially those not far from the diagonal, were yet of an appreciable size. The derived multiple regression weights (Table III) necessarily display the same features. Further, the multiple correlations are largest for the middle sets where prediction is best, and smallest for the extreme sets where prediction has to be based solely or chiefly on either the more complex or the less complex sets, not on both.

¹ It was suspected that the correlation matrix might have been influenced by the speed factor, since testees who were too slow to complete an earlier set might not attempt the later sets at all. Correlations were therefore recalculated excluding persons who had omitted any set. The resulting matrix behaved very similarly to the original, but cannot be published without indicating the general level of the testees.

TABLE III. MULTIPLE CORRELATION AND REGRESSION WEIGHTS OF EACH SET ON THE OTHERS

Predicted Set	Beta-Weights of Predictor Sets					Multiple Correlation
	A	B	C	D	E	
A	—	.350	.212	.117	.054	0.619
B	.289	—	.381	.130	.044	0.700
C	.152	.331	—	.393	.024	0.747
D	.090	.121	.421	—	.278	0.725
E	.063	.063	.039	.425	—	0.526

Apparently, the Progressive Matrices do not form a perfect Simplex, though, on the whole, Simplex theory gives a close approximation to their structure. No over-all tests of significance were available, but in view of the large number of cases tested, sampling errors cannot have been large.

To compare Guttman's Simplex approach with the traditional method of factor analysis, an analysis by principal components was also carried out. The loadings of all five principal components were computed (Table IV); and the components were found to behave exactly as the latent vectors of a Simplex matrix should.¹

TABLE IV. LOADINGS FOR THE PRINCIPAL COMPONENTS OF THE SETS

Component	Sets					Per Cent. of Total Variance
	A	B	C	D	E	
I	.746	.809	.845	.828	.635	60
II	-.354	-.287	-.134	.198	.707	15
III	.507	-.051	-.319	-.290	.268	10
IV	-.232	.489	-.067	-.282	.138	8
V	.000	-.158	.405	-.328	.100	6

The sequences of loadings for each principal component are illustrated in Fig. 1, where freehand curves have been drawn through the points to indicate their order. The loadings of the first component were all positive. Those of the second component moved from maximum negative to maximum positive, crossing zero between sets C and D. Those of the third moved from positive to a negative minimum between sets C and D, and then turned positive again. Those of the fourth moved from negative to positive, then back to negative and to positive again: the signs of the last seem to change between every two adjoining sets. This oscillation of the loadings is precisely what is entailed by the Simplex correlation structure, where the direction cosines of the first latent vector are all positive, and each successive latent vector has one more change of sign in the sequence of its direction cosines.

¹ In a previous number of this *Journal* (3), Miss Keir gives a similar table of loadings obtained by a factor analysis of the items in sets B, C, D, and E. Allowing for the interchange of difficulty in sets C and D with her subjects, the sign pattern is much the same as that shown above in Table IV.

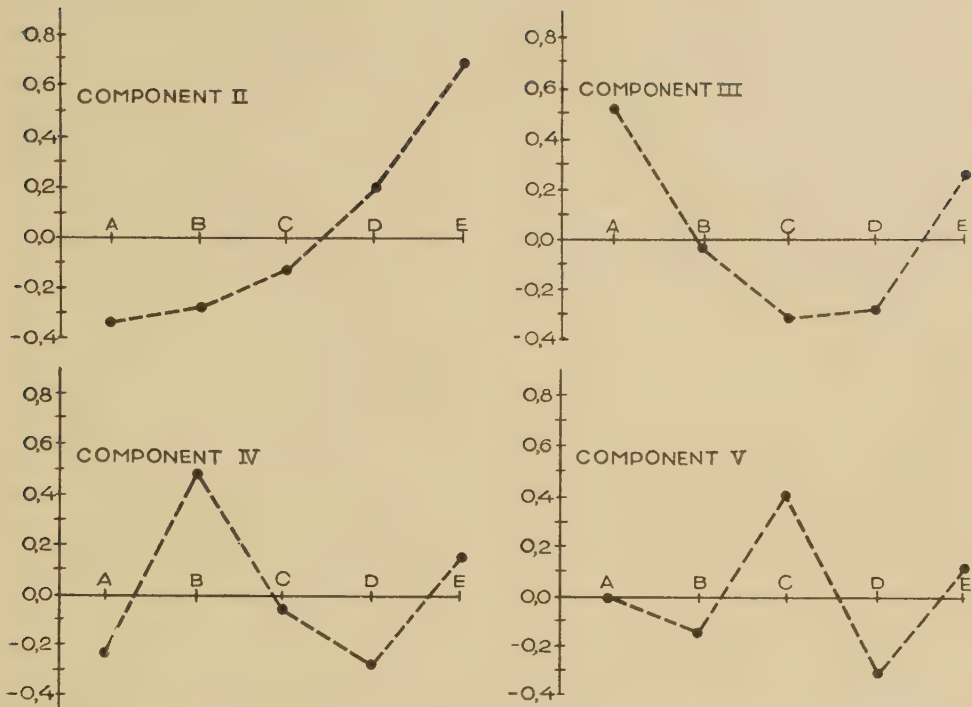


FIG. 1.—The Principal Component Loadings as Functions of the Rank Order of the Sets.
Based on Table IV

Note.—The oscillatory functions of the difference equations of the perfect Simplex are not the principal component loadings themselves, but the loadings multiplied by certain positive weights. The oscillations, however, depend essentially on sign changes, and thus are not affected by these weights, so here the loadings have been plotted directly. The first principal component, as in Table IV, has all its loadings positive; it therefore shows no sign change, and consequently has not been included in the above diagram.

IV. THE VALUE OF SIMPLEX ANALYSIS AND ITS USE IN PREDICTION

Since we lack any clear hypothesis of meaningful orthogonal factors, an analysis into principal components yields no explanation for the structure of a test battery, but only a mathematical solution.¹ On the other hand, the hypothesis of an order of

¹ Professor Guttman has drawn my attention to the resemblance of his Simplex principal component to those of his perfect scale of dichotomies (5), and also to the differences between them. He originally established the oscillatory law (exemplified in Fig. 1) for scalogram analysis and for the principal components of scale scores as well as of scale weights. For attitudinal data, psychological meaning has been found for some of the principal components of scale scores, such as 'intensity,' 'closure,' and 'involution' (8, 11). His present theory of the Simplex is for quantitative variables, not qualitative dichotomies; and the oscillatory law does *not* hold for principal component scores. The reason is that principal components are always orthogonal; and for quantitative variables—in particular those with a normal multivariate distribution—orthogonality means complete statistical independence. Not so for principal component weights, even in the quantitative case, as Simplex theory shows and our present example verifies. There is as yet no direct psychological interpretation of the principal components of a Simplex of mental tests such as is now available for attitudinal data.

complexity of tests involving cumulative functions of elemental components of ability enables Simplex analysis to explain this structure. The special value of Simplex analysis therefore lies in providing and testing a clear hypothesis of test structure.

Apart from the theoretical interest in ascertaining the structure of the sets in the Progressive Matrices test, Simplex analysis also helps us in the practical applications. For predictive use in screening or selection we require to know the ability of the test to assess some particular level of complexity. If a set exists at just that level, it, and it alone, will be maximally valid for our purposes. Other sets will either involve more, or fewer, elemental components, and will therefore contribute nothing further to validity. If no single set has that precise level of complexity, a combination of the next more complex and the next less complex sets will provide the best predictor of the required order ; and the inclusion of more distant sets will in no way improve the validity.

The order of complexity required to predict any one criterion can be ascertained from the vector giving the validities of the Simplex. The correlations should increase up to the required order, i.e., so long as the elemental components in the successive sets form an increasing part of those involved in the criterion ; they should then decrease, as more superfluous components become included in the test-items. Consequently, by locating the maximal correlations in the vector of validities, we are able to identify the approximate order of complexity possessed by our criterion, and can choose the most appropriate sets to be combined in order to ensure maximal validity.

REFERENCES

1. Burt, C. (1921). *Mental and Scholastic Tests*. London : Staples Press.
2. Raven, J. C. (1938). *Perceptual Tests of Intelligence* (M.Sc. Thesis, University of London Library).
3. Keir, G. (1949). 'Progressive Matrices as Applied to School Children.' *Brit. J. Psychol., Stat. Sect.*, II, 140-50.
4. Burt, C. (1950). 'The Theory of Mental Testing,' *Third Yearbook of Education*. London : Evans Bros.
5. Guttman, L. (1950). *Ap. Stouffer, S. A. et al., Measurement and Prediction*. Princeton : University Press.
6. Banks, C., and Sinha, U. (1951). 'An Item Analysis of the Progressive Matrices Test.' *Brit. J. Psychol., Stat. Sect.*, IV, 91-4.
7. Guttman, L. (1953). 'Two New Approaches to Factor Analysis,' *ap. Officer of Naval Research, Report on Project Nonr-731(00)*. Washington, D.C., Government Publications.
8. Guttman, L. (1953). 'A Note on Cyril Burt's Factorial Analysis of Qualitative Data.' *Brit. J. Stat. Psychol.*, VI, 1-4.
9. Burt, C. (1953). 'Scale Analysis and Factor Analysis,' *ibid.*, VI, 5-23.
10. Guttman, L. (in the press). 'The Radex : A New Approach to Factor Analysis,' *ap. Lazarsfeld, P., Mathematical Thinking in the Social Sciences*. New York : Columbia University Press.
11. Guttman, L. (in the press). 'The Principal Components of Scalable Attitudes,' *ibid.*

THE SIGN PATTERN OF FACTOR-MATRICES

A Note on a Problem raised by Mr. Gabriel

By CYRIL BURT

University College, London

I. *Progressive Sign-Changes.* II. *Theoretical Examples.* III. *Empirical Examples.* IV. *Summary and Conclusions.*

I. PROGRESSIVE SIGN-CHANGES

The General Problem. In his discussion of the 'Simplex structure' underlying the progressive matrices test, Mr. Gabriel has called attention to the peculiar oscillatory pattern of signs characterizing the factor-loadings that he has obtained with the method of 'principal components' (16, Fig. 1). What is the explanation? Although in the case of 'attitudinal data' a plausible reason for such peculiarities has recently been put forward by Professor Guttman, there is, we are told, in the case of data derived from mental tests, "as yet no direct psychological interpretation." We are still "lacking any clear hypothesis of meaningful orthogonal factors"; and consequently we can expect "no explanation of the structure of a test battery from a principal component analysis." The chief object of this note is to suggest a simple derivation, which in theory seems applicable to any kind of data, and should therefore furnish a "direct psychological interpretation" when the data are obtained from mental tests.

The view I am putting forward has an important practical bearing. In factorial studies submitted to this *Journal*, writers often set out their tests in an order which is quite irrelevant to the question they wish to decide: sometimes, for instance, they merely leave them in the order in which they were administered. As a consequence, it is by no means easy to discern the structure of the factor-matrix as it is printed. Take, for example, what is the largest and perhaps the most interesting of factor-matrices we have published (Table III, this *Journal*, I, p. 120), and then note how much more revealing the results become when a later writer re-arranges the variables so as to bring out the pattern of signs (Table II, *ibid.*, p. 210). With most tables of this kind, once the tests have been grouped according to their signs, the interpretation of the first two or three bipolar factors strikes the reader at a glance. To interpret the later factors is more difficult: they have to be inspected, not singly, but in sets or types. And, probably because such points can be explained so much more readily by word of mouth, there seems to be little or no published account of the principles involved.

Solution. Briefly the explanation I wish to suggest is as follows. Factor-measurements are derived from observed measurements by a reiterated process of addition and subtraction. For these two operations the conventional symbols are the plus and minus signs. Hence, if the bases of classification indicated by successive factors are mutually independent, then, as the classificatory subdivisions get more and more numerous, so the sign-changes in each factor-column must become more and more frequent. The principle is by no means new. It has cropped up incidentally in several earlier discussions, and may be reached by at least three different lines of approach—operational, statistical, and logical. Yet all turn on the same fundamental conception.

(a) *Operational.* Primarily a factor is specified, not by the factor-saturations, which are merely weights for reconstructing the test-measurements from the factors, but by the factor-measurements. And the basic assumption governing all our analyses is that the figures observed (let us call them the 'test-measurements') are in theory built up from the

several factor-measurements by a process of successive addition. It follows that the factor-measurements may be derived from the test-measurements by the inverse process, that is, by a process of successive subtraction. At each stage the particular measurements to be subtracted are found by averaging, that is, by weighted addition; but for an ordinary or simple average the 'weights' are all numerically equal, namely, $1/n$. Now at every succeeding step the subtraction must introduce one extra change in sign. It would be tedious to offer a verbal demonstration. The reader can quickly convince himself that this is so, if he attempts to work out a short example by the method of simple averaging (cf. this *Journal*, I, p. 5, 'Averaging Successive Deviations').

It may also be shown that precisely the same factor-measurements can be calculated much more speedily by applying a matrix of weights, not to the successive deviations or residuals, but to the initial trait-measurements. In general, however, the weights required will still exhibit the same progressive increase in the number of sign-changes as was discernible when applying the method of successive averaging. Thus factor-weights will, almost inevitably, conform (or at least tend to conform) to what may be called the *law of increasing sign reversals*.

(b) *Statistical*. In practice, instead of working with the actual test-measurements, the statistician shortens the process still further by basing his calculations on the test correlations. Now, when factors are computed either by 'simple summation' (centroid method) or by 'weighted summation' (method of principal components), the factor-measurements themselves will be uncorrelated, i.e., 'statistically independent' (4, p. 50) or, as I should prefer to say, orthogonal; and the weights will have, not only different signs, but also different numerical values. With empirical data this will often introduce appreciable irregularities into the sign pattern. If we adopt the method of principal components, then each column of factor-saturations (the regressions of the tests on the factors) will be proportional to the corresponding row of factor-weights (the regressions of the factors on the tests). They must therefore exhibit the same pattern of signs, whether regular or irregular. And to the experienced factorist both the regularities and irregularities will yield considerable insight into the data he is analysing, even without any further rotation or analysis.

The method of principal components has one further peculiarity: the columns of factor-saturations, as well as the rows of factor-measurements, are mutually orthogonal. This means that the product-sum of any two columns must be exactly zero (it is often said that they are 'uncorrelated,' but the phrase is only true if it refers to the 'unadjusted correlation'). With 'simple summation' orthogonality is only approximately achieved: in that case the condition imposed is that every column of saturations (except the first) must add up to zero just as it stands, and when weighted by the sign pattern of every other column. Now consider the effect of such conditions on the sign pattern. For simplicity, let us take an imaginary case in which the saturations in each column are all equal (without necessarily being equal to those in the other columns). It will be obvious, when we apply the condition factor by factor, that a zero product-sum between every pair of columns can occur only if the pattern of signs obeys the law of increasing sign reversals just described.

Examine, for instance, the sign pattern in the factor-matrix for eight tests set out in Table IB, or better, let the reader deduce for himself the arrangement of plus and minus signs required to fulfil the conditions imposed. He will find that, if he tries to keep similar signs together in each column, there must be one more change of signs, and only one, in each succeeding column (and therefore in each succeeding row). With eight tests, exactly eight factor-columns, and no more, can be constructed to comply with this principle.

In an actual factor-matrix, the saturations will not be equal. But, unless the amounts differ widely, the same kind of pattern will tend to appear if orthogonality is to be preserved. A striking illustration is to be found in the values for orthogonal polynomials of degree 1, 2, 3, . . . , n , where the same law of sign-change is deducible from the way orthogonality is secured.¹ As will be seen from Table IV in Mr. Gabriel's paper, these are the proportions to which the 'loadings' approximate when we are factorizing what he calls a 'Simplex structure.'

¹ Cf. Fisher, R. A., and Yates, F., *Statistical Tables* (1943), Table XXIII, for $n = 3$ to $n = 52$. I have shown how these values arise in the principal components solution for certain data, and how the sign values are determined by the general algebraic formula (15, pp. 13, 21f.).

Dr. Gabriel states that "the oscillatory law does not hold for principal component scores." But, since the 'scores' (i.e., the factor-measurements) are also orthogonal, a similar tendency must be at work. This, indeed, follows from what I have called the 'reciprocity principle.' But, in ordinary investigations, the fact that the number of persons so greatly exceeds the number of factors is apt to obscure the regularity of the pattern in the resulting scores.

(c) *Logical.* The phenomenon I have described is not a mere mathematical curiosity, having no special implications for the psychologist. In an earlier paper (5) I argued that the sign pattern of a bipolar factor is in effect a dichotomous sorting machine; as I endeavoured to show, the vectors of weighting coefficients derived by the method of principal components have all the properties of what in quantum theory are technically termed 'a spectral set of selective operators.' Eddington, in explaining the nature of a 'selective operator,' reminds his readers of the answer to the schoolboy's conundrum, "how do you catch lions in a desert?" The answer is: "Take a large sieve; sift out the sand; and what remains will be lions."

Carrying this principle a little further, in order to demonstrate the effects of a succession of such factors, I use a simple model based on the well-known 'Pearson-Galton Binomial Apparatus' (Yule, 1910 ed., Fig. 48; Yule and Kendall, Fig. 10. 2). The material to be sorted flows down a set of branching channels; and at the entry to each channel a sieve is inserted. The whole machine thus consists of a pyramid of sieves. The top sieve has holes of $\frac{1}{2}$ inch; the two below have holes of $\frac{1}{4}$ inch and $\frac{3}{4}$ inch; the four below, holes of $\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{8}$, and $\frac{7}{8}$ inches; and so on. Globular beads of various sizes are poured into the funnel. The first sieve separates them into two main groups; the tinier beads pass through it and roll down to the $\frac{1}{4}$ -inch sieve, while the larger are held back and tipped into the $\frac{3}{4}$ -inch sieve. Each of these two lower sieves again sifts what it receives into two smaller groups—four groups in all. And a similar twofold segregation is repeated on every subgroup at each successive stage.

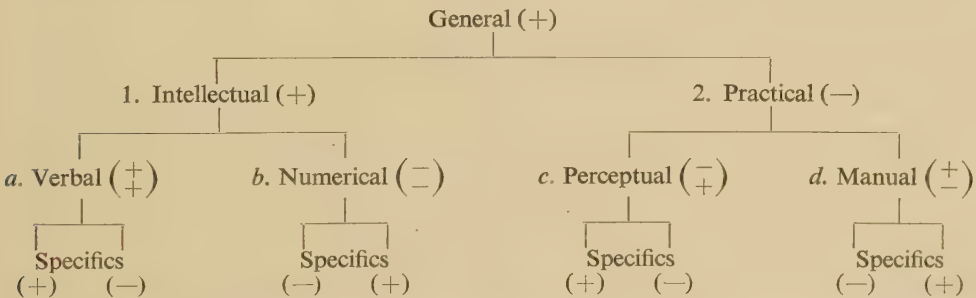


FIG. 1.—Tree to Illustrate Hierarchical Classification of Eight Traits by Factor Analysis.

Note.—In a complete analysis there would be *four* signs to each specific.

In the factor-matrix the general factor corresponds to the funnel, and the first, second, and n th bipolar to the first, second, and n th row of sieves. At any one stage their plus and minus signs subdivide into two contrasted classes whatever is contained in the positive or the negative division of the preceding factor. Since at each new level and with each new factor the number of dichotomies is doubled, the process results in the tree-like scheme of classification which the formal logician terms 'hierarchical' (cf. Fig. 1).

This conception of bipolar factors as 'bipartite selective operators' should, I think, provide what Mr. Gabriel is looking for—an explanation that will give intelligible meaning to a set of orthogonal factors. So long as we insist on regarding factors as mental abilities or faculties, then, as he rightly observes, any direct psychological interpretation of the factors produced by summational procedures will be difficult to suggest. Thurstone, for example, who apparently holds that factors must always be abilities (4, p. 53), declares that neither the centroid method nor the method of principal axes yields "psychologically meaningful situations." But if, as I have always contended, the factors so obtained represent, not

concrete abilities but abstract principles of classification, then the sign pattern is seen at once to serve a most important purpose: it classifies, so far as they admit of such a classification, the initial traits or items in accordance with a dichotomous hierarchical scheme.

Having obtained a formal classification of this kind, there is no reason why we should not, as logicians usually recommend, change it to some more concrete classification, appropriate to the material in question. In certain enquiries we may rightly prefer to substitute factors that can be directly identified with mental abilities. Now an ability for x is by definition a dispositional property that *facilitates* doing x , i.e., it denotes a positive and never a negative tendency. Hence we shall be compelled to seek factors with positive saturations only. This can readily be done by converting the bipolar factor-matrix into a group factor-matrix by a suitable linear transformation. But it is plain that, if we are to derive any empirical information from the results of our analysis, this transformation must, so far as possible, preserve the original hierarchical scheme. A rotation which allows us to impose at will almost any preconceived scheme of our own can prove little or nothing: it is a method, not of fact-finding, but of wish-fulfilment.

However, before we go on to study the effects of such rotations, it will be well to set down formal definitions for the several terms that I have hitherto been using in a loose and self-explanatory fashion.

Definition of a Dichotomous Hierarchy. To formulate more precisely the mathematical model that I have been trying to suggest, let us start by considering the simplest conceivable case. The method of classification I have just described corresponds to what in formal logic is called 'division by dichotomy.' According to the usual definition, this is "a classificatory procedure such that at each stage the division is into two, and only two classes, distinguished by the fact that their members do, or do not, possess a given differentia." The postulates on which it rests are the 'principle of contradiction' and the 'principle of excluded middle'; and, so far as these are fulfilled, the procedure necessarily satisfies all the rules for a valid logical division. Hence it "possesses the great advantage that it is always formally valid."

On the other hand, "this formal division does not of itself ensure that the resulting classes exist," i.e., contain actual members.¹ Where such classes exist, the classification gives rise to what are called 'taxonomic groups.' These are the 'groups' that we want our group factors to designate. They may be defined² by saying that "they are such that, if we take any two, then either they have no member in common, or else one group is wholly contained in the other" (e.g., if, following Bentham, we divide all living organisms into (a) the animal kingdom and (b) the vegetable kingdom, and define vegetable or plant to include all organisms that are not animals, but nothing besides, then it is impossible that anything should be a member of both kingdoms at once; at the same time, we can say that every animal will be a living organism, and so will every plant).

The relation which I have designated 'containing' will be asymmetrical and (in general) one-many; and a relation of this kind generates a 'hierarchical arrangement.' This brings us to the term that has been so confusingly used in factorial discussions. By a 'hierarchical arrangement' (R_n) I understand "a *divergent* order progressively generated by an asymmetrical, *one-many* relation," much as a *serial* order is generated by an asymmetrical, *one-one* relation. And by a 'hierarchy' I mean "the whole class of terms thus arranged, including both the single initial referent (here the *summum genus* which contains all the other classes) and the converse domain of R_n which forms the *relata*." In such a hierarchy "the n th 'level' may be defined as the class of all terms to which the initial referent has the relation R^n ."³ In the simplest possible case, in which R is a one-two relation, and the subclasses on any given level are not only mutually exclusive but also collectively exhaustive (as is

¹ The quotations are from Stebbing, L. S., *A Modern Introduction to Logic*, pp. 436f.

² For a discussion of the formal characteristics of 'taxonomic groups' in the biological sciences, see Woodger, J. H., *The Axiomatic Method in Biology* (1937), pp. 131f., from which I have taken my definition, with some slight verbal changes.

³ These definitions are quoted from an earlier article in this *Journal* (II, p. 63). As there noted, the taxonomic relation—that of 'containing' or 'class-inclusion'—is by no means the only relation that may generate a hierarchy; and in many enquiries the ultimate interpretation of a hierarchy may depend on discovering, from non-statistical studies, some analogous but more concrete relation which will explain the phenomena observed.

ensured when the one differentia is the contradictory of the other), we can state that every member in one of the wider classes will be found among the members of one or other of the two subclasses which it contains.

Expressing such a scheme in factorial terms, we see that the group factor-matrix which specifies it will comprise first a 'generic' or 'basic' factor, covering all the traits and thus describing the 'universe' or 'genus' that we are proposing to classify; then, since the classification is dichotomous, this 'genus' will be divided into two 'species,' indicated by two broad group factors: and, if the sample of traits or variables is fairly representative, each of these species will include approximately an equal number of such variables. Then each of the two species will be subdivided into a couple of subspecies, indicated by narrower group factors; but, as a rule, the differentia or 'fundamentum divisionis' which effects the subdivision will not be the same for both species (e.g., Bentham's celebrated 'bifurcate system' went on to divide the vegetable kingdom into (ai) flowering and (aii) flowerless, and the animal kingdom into (bi) vertebrates and (bii) invertebrates: *Essai sur le Classification*, 1823, pp. 107f.). At the next stage we may suppose each subspecies will be similarly subdivided into sub-subspecies. And so we shall continue until we reach the *infimae species* which in factorial work will be described by a single variable only.

As I have observed, we cannot be sure in advance that a given set of data will fulfil the conditions necessary for a systematic classification, much less what shape that classification will take. Moreover, many if not most of the characteristics that concern us in psychology will not be simple qualities, such as are either wholly present or wholly absent; they will be tendencies which may manifest themselves in varying degrees, and they will do so as complex patterns rather than as single attributes. Hence in psychology classification becomes a statistical rather than a merely logical problem. We may therefore regard factor analysis as a convenient technique for progressively screening the data under investigation in order to determine (a) how far they actually fit into a hierarchical classification, and (b) what is the concrete nature of the scheme when we have discovered it.¹

II. THEORETICAL EXAMPLES

Group Factor Schemes. For the arithmetical implications of this formulation, let us look first at a purely artificial illustration. For a perfect example of the simplified type I have just described, the number of traits must be an integral power of two. Let us therefore take a series of $2^3 = 8$ traits or tests. In such a case a complete hierarchical structure of the kind I have outlined would yield a scheme of subdivided group factors, like that set out in Table IA. What particular traits, tests, or other variables we actually choose will depend on the nature of the material and of the hypothesis we wish to verify or refute.

¹ Critics like Dr. Chambers and Mr. Kenneth Richmond will perhaps object that I have "neither stated nor defended [my] basic assumptions." Factorists, they say, "in their architectural zeal" invariably ignore the "foundations on which their numerical edifices stand" (Zangwill, O., *Introduction to Modern Psychology*, 1950, pp. 142, 160). Hence, however correct the mathematics may be, the "figures turned out by the factorial sausage machine" can claim neither theoretical value nor practical utility. Briefly my 'basic assumptions' are three: (i) that in psychology, as in other sciences, an essential preliminary is to classify the phenomena with which we are concerned; (ii) that the ideal classification is a hierarchical classification; and (iii) that, from a formal standpoint, the most general method of achieving such a classification is to proceed by dichotomy.

The shortest way to defend such assumptions will be to quote an authoritative statement in support of each one, and refer the reader to the fuller discussion in references cited. "Science," says Whitehead, "is organized thought; hence the scientist's first step must be to classify. . . . But unless you can progress from classification to mathematics, your reasoning will not take you far" (*Science and the Modern World*, pp. 43, 244; cf. *Concept of Nature*, ch. IV). "The ideal of classification," says Miss Stebbing, "is so to arrange classes that their relations may be exhibited in accordance with the principles of hierarchical order" (*Introduction to Modern Logic*, p. 439). "A bifurcate classification," says Jevons, "is the inevitable and only scheme that is logically perfect; all other arrangements can be made to correspond to a bifurcate arrangement, with the implication that some of the classes may have no representatives" (*Principles of Science*, p. 694). "The fundamental assumption which is involved in an all-embracing classificatory system is that the particular universe of things presented to us must be capable of being arranged by a logical division, in which we proceed from the highest and most general concept by way of progressive determination through opposite characteristics" (Sigwart, C., *Logic*, II, p. 512).

TABLE I. HIERARCHICAL FACTOR SCHEME FOR EIGHT ITEMS
(1) With Unreduced Self-Correlations

Item	A. Group Factors								B. Bipolar Factors							
	I	II	III	IV	V	VI	VII	VIII-XV	I'	II'	III'	IV'	V'	VI'	VII'	VIII'
1	.5	.5		.5				.5	.68	+.47	+.31	+.31	+.18	+.18	+.18	+.18
2	.5	.5		.5				etc.	.68	+.47	+.31	+.31	-.18	-.18	-.18	-.18
3	.5	.5			.5				.68	+.47	-.31	-.31	-.18	-.18	+.18	+.18
4	.5	.5			.5				.68	+.47	-.31	-.31	+.18	+.18	-.18	-.18
5	.5		.5			.5			.68	-.47	-.31	+.31	+.18	-.18	-.18	+.18
6	.5		.5			.5			.68	-.47	-.31	+.31	+.18	+.18	-.18	-.18
7	.5		.5				.5		.68	-.47	+.31	-.31	-.18	+.18	-.18	+.18
8	.5		.5				.5		.68	-.47	+.31	-.31	+.18	-.18	+.18	-.18
Exact Values			..		..		..		$\sqrt{\frac{15}{32}}$	$\sqrt{\frac{7}{32}}$	$\sqrt{\frac{3}{32}}$	$\sqrt{\frac{3}{32}}$	$\sqrt{\frac{1}{32}}$	$\sqrt{\frac{1}{32}}$	$\sqrt{\frac{1}{32}}$	$\sqrt{\frac{1}{32}}$
Contribution to variance (per cent.)			..		..		..		46.9	21.9	9.4	9.4	3.1	3.1	3.1	3.1
Number of sign-changes			..		..		..		0	1	2	3	4	5	6	7

Note.—In Table IA eight specific factors are needed to complete the scheme ; but to save space all except the first have been omitted. With these the variance for each item adds up to unity.

In the majority of factorial enquiries the assumed differences between the alleged factors are qualitative. For example, the investigator may wish to discover whether, over and above what is known as 'general ability' or 'intelligence,' there may not be more specialized factors corresponding to the different modes of cognitive activity—sensory, perceptual, associative, and the like. One of the commonest schemes reported is that illustrated in Fig. 1 (the reason for the double signs in the third row is explained in the paragraph on 'subdivided factors'). There are several minor variations of the scheme, especially in regard to the more specialized traits ; for examples, see this *Journal*, II, pp. 84, 87, and Vernon (12).

But in other enquiries the processes tested are presumed to differ, not in quality, but merely in degree of complexity or difficulty. This was Spearman's interpretation of the so-called 'hierarchy of intelligences' (3, p. 69), since he maintained that the alleged qualitative differences did not in fact exist. Less controversial examples may be found in experiments on the span of attention or of immediate memory with uniform material, where the sole difference between the various sub-tests lies in the number of digits, letters, or syllables to be perceived or reproduced (cf. this *Journal*, VI, Table I, p. 7). A similar principle was adopted in the earliest versions of the progressive matrices test : by constructing square patterns of increasing size and complexity, a graded series of non-verbal 'analogies' could be built up with the component items increasing in difficulty by fairly uniform steps.¹

In cases of this latter type, when a group factor analysis is carried out, factor I will usually represent the general level of the tests : factor II the easier half and factor III the harder half ; factor IV will represent the easiest quarter, and so on. Thus, in theory, we should again have a factor-pattern resembling that shown in Table IA, where we may suppose the eight items to have been arranged in order of decreasing ease and therefore of increasing complexity. (This way of dealing with graded characteristics has unexpected uses : it is embodied in the construction of many modern electronic computers, where measurable quantities are expressed on a binary scale of notation, i.e., by a succession of disjunctions : cf. also this *Journal*, IV, pp. 196f.)

Analysis by Principal Components. The essential differences between group factor-patterns and bipolar patterns can be seen most clearly if we again take the case where the

¹ Cf. Burt, C., *The Backward Child*, 1937, pp. 55-6. The test is there termed 'serial analogies' ; but the scheme of construction is set out as an $n \times n$ matrix.

group factor-saturations are all equal. Let us therefore construct a correlation table from the saturations given in Table IA, and then (keeping the self-correlations at unity) carry out a complete analysis of the results by Pearson's method of principal components. Since there are eight items we shall obtain eight factors. The new factor-matrix is set out in Table IB. The figures are no longer the same throughout. And, since the procedure seeks at each stage what Pearson called 'the lines of closest fit,' the factors are extracted in order of diminishing variance. The factors are tabulated in this order¹; and it will be noted that the number of sign-changes increases step by step from 0 to 7.

The meaning of the bipolar factors is obvious directly we compare Table IB with Table IA. The contrast between group factors II and III is now represented by the contrast in signs within bipolar factor II'. Similarly, if there were only four items, the contrast between group factors IV and V would be represented by the contrast in signs within bipolar factor III'. Here, however, there are eight; and, since by the nature of the method factor III', like all the rest, is common to *all* the items, the same contrast of signs expresses the difference between group factors VI and VII, which enter only into the last four items. But this introduces cross-correlations which have no foundation in the data: thus, the bipolar factor III' yields a positive cross-correlation between items 1 and 8: but, after the first two components have been removed, there is no such correlation; the residual is zero. Similarly with other cross-correlations (between 2 and 7, 3 and 6, etc.). To neutralize these unwanted correlations, a further bipolar factor is needed, with the same numerical values, but with signs reversed for the last four items, namely, bipolar factor IV' (cf. Fig. 1). Indeed, with empirical data, the fact that the group factors are 'subdivided' is generally revealed by the appearance of zero (or near-zero) residuals after the first two factors have been extracted by the summational procedure (cf. this *Journal*, II, p. 47). Exactly the same peculiarities recur after the first four bipolar factors have been extracted; and exactly the same reasons explain why the last four bipolar factors (V', VI', VII', and VIII') have the same numerical values and the same subdivisions, but opposite signs in some of the subdivisions (in Fig. 1 only one row of signs has been inserted in the bottom line).

But what has happened to the eight specifics (VIII to XV)? These have been incorporated into the eight *common* factors furnished by the principal component analysis. It will be seen that in Table IB each item has its own 'specific' sign pattern, i.e., a row of signs which is peculiar to itself. By this means the separate specific factors which appear in the group factor analysis are accounted for in the principal component analysis without introducing separate specifics as such. As a result, the total number of factors is reduced from fifteen to eight—eight being the rank of the complete correlation matrix.

Algebraic Formulation. The reader will have no difficulty in deducing a rule for determining the number of factors of each different type in simplified cases like the above, where the group factor-saturations are all the same. Thus, with $n = 2^m$ tests, the number of different factors having 2^{m-r} non-zero saturations will be 2^r : e.g., the number of factors having $2^m = n$ saturations will be $2^0 = 1$ (i.e., there will be only one general factor); and the number of factors having $2^0 = 1$ saturation will be $2^m = n$ (i.e., there will be n specifics). In the bipolar analysis, we may class together all factors contributing the same amount to the total variance. Then, since $2^m = 2^0 + 2^0 + 2^1 + 2^2 + 2^3 + \dots + 2^{m-1}$, the number of factors having 0, 1, 2, 3, 4, ..., 2^{m-1} sign-changes will be respectively 1, 1, 2, 4, 8, ..., 2^{m-1} . And the contributions of these various types to the variance of each test will be:

$$\frac{2^{m+1} - 1}{(m+1)2^m}, \frac{2^m - 1}{(m+1)2^m}, \dots, \frac{2^1 - 1}{(m+1)2^m}, \text{ respectively.}$$

The total contribution will therefore be:

$$\begin{aligned} 2^0 \frac{2^{m+1} - 1}{(m+1)2^m} + 2^0 \frac{2^m - 1}{(m+1)2^m} + 2^1 \frac{2^{m-1} - 1}{(m+1)2^m} + 2^2 \frac{2^{m-2} - 1}{(m+1)2^m} + \dots + 2^{m-1} \frac{(2^1 - 1)}{(m+1)2^m} \\ = \frac{(2^{m+1} - 2^m) + m 2^m}{(m+1)2^m} = 1. \end{aligned}$$

¹ With unspecified items, and with equal saturations, we have a certain amount of freedom in drawing up the arrangement; but with a well-planned empirical research, the freedom will be very narrowly confined, and the arrangement may even be uniquely determined.

The Ideal Rotation Matrix. The numerical relations between the group factors and the summational factors will become plain if we examine the transformation matrix required to rotate the matrix of group factors to the matrix of principal components. As the space required to print matrices of 8×15 would be excessive, let us take a table with only four items. This will be sufficient to exhibit the underlying principles. The transformation matrix we need will be a section of an orthogonal matrix, as may readily be verified with the table below ; its elements can be easily computed by the method described in *Factors of the Mind*, pp. 304f. (see Table II, top panel).

TABLE II. ROTATION OF GROUP FACTORS TO PRINCIPAL COMPONENTS

Item								Factor	I	II	III	IV
	I'	II'	III'	IV'	V'	VI'	VII'					
								I'	$\frac{2}{\sqrt{7}}$	0	0	0
								II'	$\frac{1}{\sqrt{7}}$	$+\frac{1}{\sqrt{3}}$	0	0
								III'	$\frac{1}{\sqrt{7}}$	$-\frac{1}{\sqrt{3}}$	0	0
								IV'	$\frac{1}{2\sqrt{7}}$	$+\frac{1}{2\sqrt{3}}$	$+\frac{1}{2}$	$+\frac{1}{2}$
								V'	$\frac{1}{2\sqrt{7}}$	$+\frac{1}{2\sqrt{3}}$	$-\frac{1}{2}$	$-\frac{1}{2}$
								VI'	$\frac{1}{2\sqrt{7}}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{2}$	$+\frac{1}{2}$
								VII'	$\frac{1}{2\sqrt{7}}$	$-\frac{1}{2\sqrt{3}}$	$+\frac{1}{2}$	$-\frac{1}{2}$
1	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$		$\frac{1}{\sqrt{3}}$					$\frac{7}{2\sqrt{3}}$	$+\frac{1}{2}$	$+\frac{1}{2\sqrt{3}}$	$+\frac{1}{2\sqrt{3}}$
2	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$			$\frac{1}{\sqrt{3}}$				$\frac{7}{2\sqrt{3}}$	$+\frac{1}{2}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$
3	$\frac{1}{\sqrt{3}}$		$\frac{1}{\sqrt{3}}$			$\frac{1}{\sqrt{3}}$			$\frac{7}{2\sqrt{3}}$	$-\frac{1}{2}$	$-\frac{1}{2\sqrt{3}}$	$+\frac{1}{2\sqrt{3}}$
4	$\frac{1}{\sqrt{3}}$		$\frac{1}{\sqrt{3}}$				$\frac{1}{\sqrt{3}}$		$\frac{7}{2\sqrt{3}}$	$-\frac{1}{2}$	$+\frac{1}{2\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$

As the figures indicate, the general factor in the bipolar analysis absorbs a portion from *all* the factors in the group factor-matrix : it incorporates them in the proportions of four from the 'basic' factor, two from the group factors, and one from the specific. (This incidentally is a point overlooked by many factorists : I have seen papers which argue that because the general factor obtained by a summational or centroid analysis increases from year to year, therefore the child's special abilities diminish ; whereas the increase in the variance of the general factor might easily be due to an increase, not in the basic factor, but in the group factors.) The first bipolar (factor II) expresses the difference between the two main group factors (factors II' and III'), but it also incorporates part of the difference between the specifics. The last two bipolars (factors III' and IV') are based solely on the specifics, which they convert to common factors. But, as we have just seen, they do not absorb *all* the factor variance contributed by the specifics ; half of it is spread over all the earlier factors.

The transpose of the rotation matrix will of course transform the bipolar matrix back to group factors.

As we have taken the simplest conceivable scheme for our group factors, the matrix shown in Table II will also give us the simplest conceivable scheme for a 4×4 rotation matrix. However, the actual size of the weighting coefficients never greatly matters. Hence a matrix of this type will yield quite good results with *any* bipolar matrix of the same order and sign pattern, even when the saturations differ appreciably from the figures we have used. Thus stock rotation matrices may be constructed in advance for any given number of traits, and will furnish a useful start in transforming an empirical bipolar matrix to group factor form. If minute exactitude is not important, the group factor matrix so computed can then be readily adjusted by two or three minor graphical rotations, chosen so as to reduce the smaller or non-significant saturations as nearly as possible to zero. In this way much of the time usually spent on rotations can be saved. Similar stock rotation matrices can be drawn up for analyses based on a 'reduced' correlation table (cf. this *Journal*, I, p. 8).

Completely Overlapping Group Factors. In his article, Mr. Gabriel is concerned primarily with composite tests or test-scales which, like the Progressive Matrices or the Binet Scale, can be regarded as based on what I have called the principle of 'completely overlapping group factors.' As we have seen, if the scale of tests were perfect, all differences of quality would have been eliminated. Every item would measure the same fundamental characteristic, namely, intelligence; but they would measure it at successive levels of difficulty, e.g., levels like those which Binet described as successive 'mental ages.' The abilities required to answer items of different levels may be represented by different group factors, and the factors will then form a triangular factor-matrix. Such factors, however, can no longer be identified with 'special aptitudes,' i.e., aptitudes for qualitatively different types of mental process: they will be 'difficulty factors.' To solve the easiest item, the child need only possess an ability of the lowest level; if he answers the hardest, he must possess an ability of the highest level, and *a fortiori* all the other abilities. The analysis of such tests is discussed more fully in my account of 'the triangular method of factor analysis' (cf. this *Journal*, IV, pp. 43f.).

The group factors which I have described as representing successive 'levels of ability' Mr. Gabriel speaks of as the 'elemental components' of the scale. Thus, it would seem that the only difference between what he has called a 'Simplex structure' and what I called a 'structure of completely overlapping group factors' is a difference in names.

The complexity of an item may be defined by the number of group factors it contains. And, in correlation tables derived from tests or scales so constructed, when the items have been arranged in order of their complexity or difficulty, the coefficients very frequently exhibit what I have called a 'diagonal ridge.'¹ As Mr. Gabriel puts it, "in a Simplex correlation matrix . . . the highest correlations are found near the diagonal." If we proceed to factorize a table of this kind by the method of principal components, then, as I have argued (15, p. 15), "the automatic effect of the analysis is to classify the items (and the persons) according to a somewhat artificial 'hierarchical scheme': such a scheme "classifies and subclassifies persons according to levels of ability, i.e., according to the difficulty of the items they can successfully answer," and the items "according to the levels of difficulty they represent"—the first bipolar contrasting the easier half of the items with the harder, the next bipolar the easier and the harder within each section, and so on.

Reduced Self-Correlations. In most factorial researches the number of persons tested is so few that all but the first three or four factors are non-significant; and, although with self-correlations of unity a complete analysis requires these non-significant factors to be computed, the differences between most of their saturations will usually be both slight and erratic. However, little information is lost and the labour of calculation greatly lightened if the variance due to specific factors is removed at the very outset; and for these and other reasons it is now customary to work with reduced self-correlations.

Accordingly, instead of carrying out a complete analysis on Pearson's lines (thus obtaining n factors for n traits), let us now eliminate the specific factor-variance from the diagonal elements of the correlation table we have been analysing, and then factorize it afresh. The results can be readily deduced. The elimination of the specific factor from each item removes

¹ Cf. this *Journal*, IV, 1951, p. 43 and refs., and IV, p. 75. If we want one factor-matrix to reveal the nature of these completely overlapping factors (or sections), this can best be done by the triangular method of factor analysis, referred to above: indeed, as was stated in my article, the presence of a 'diagonal ridge' seems nearly always to be "a cue for attempting the triangular method."

.25 from its total variance ; and, of the amount so removed, one half is accounted for by the saturations in the last four columns of the complete bipolar analysis (Table IB), while the other half is distributed among the first four factors. Hence, from the variance of each of these four factors (I to IV) we have to subtract $1/32$. Thus the variance for the first factor will now be $\frac{15}{32} - \frac{1}{32} = \frac{7}{16} = (.6614)^2$; and similarly for the remainder. This yields the saturations shown in Table III. The reader can easily satisfy himself that these are the figures that would be obtained by applying the ordinary summational procedure to the reduced correlation matrix.

TABLE III. HIERARCHICAL FACTOR SCHEME FOR EIGHT ITEMS

(2) With Reduced Self-Correlations

B. Bipolar Factors

Item	I	II	III	IV	Variance
1	.66	.43	.25	.25	.75
2	.66	.43	.25	.25	.75
3	.66	.43	-.25	-.25	.75
4	.66	.43	-.25	-.25	.75
5	.66	-.43	-.25	.25	.75
6	.66	-.43	-.25	.25	.75
7	.66	-.43	.25	-.25	.75
8	.66	-.43	.25	-.25	.75
Exact values	$\sqrt{\frac{7}{16}}$	$\sqrt{\frac{3}{16}}$	$\sqrt{\frac{1}{16}}$	$\sqrt{\frac{1}{16}}$	—
Contribution to variance (per cent.)	43.75	18.75	6.25	6.25	—
Number of sign-changes	0	1	2	3	—

Since the tests now occur in identical pairs, the rank is reduced to four ; and four factors alone will be required. But none of these changes affects the sign pattern. Naturally, with a table empirically obtained the members in such test-pairs might differ widely from each other. As a result it is quite conceivable that one or more additional factors, fully significant, would be required ; and in certain cases these might appreciably spoil the neatness of the pattern.

III. EMPIRICAL EXAMPLES

Empirical Analyses. To make my points clear I have so far dealt with highly artificial matrices satisfying the simplest conceivable conditions. With empirical tables the relations we have discovered may, naturally enough, be somewhat blurred or altered by the peculiarities of the data. The contributions of the several group factors to the total variance will often be far from equal and (unless the factors are numerous) nearly always larger than the contributions of the specific factors. Nor as a rule will the number of tests be an exact integral power of 2 ; and the consequent derangement of the balance will tend still further to blur the sign pattern. Such disturbances most frequently affect the later factors. Thus, with a complete analysis by summational methods, the specific factors often show an increasing number of zero (or near-zero) saturations, and so appear to be represented, not by a series of common factors, but by small sets of subdivided factors, or by small triangular matrices of completely overlapping factors, or by factors which are themselves very nearly specific. In many factorial studies the tests or assessments analysed by the investigator have originally been collected for some very different purpose, and so may not contain a sufficiently wide or

sufficiently well-balanced selection of traits. In such cases we can hardly hope for conclusive evidence on more general problems. Yet, strangely enough, a serious distortion of the results is much rarer than one would expect.

In spite of all these modifications, the law of progressive sign-change seems usually to hold for those bipolar factors that do not represent specifics, and is often discernible within the submatrices representing the specifics. Moreover—and this is the point I want more particularly to emphasize—a survey of the various analyses that have been carried out with sufficient data and sufficiently rigorous methods will, I think, quickly convince the reader that the law obtains far more widely than Mr. Gabriel has supposed. (i) It is discernible in psychological data that do not conform to the requirements of 'Simplex structure.' (ii) It is even discernible with data that are not psychological at all, e.g., physical measurements. (iii) It is discernible, not only when reduced self-correlations are substituted for unity, but also when simple summation has been substituted for weighted summation.

The tendency, however, is quite needlessly obscured when the computer forsakes the strict and stereotyped routine of the summational procedures. To save time many factorists now follow Thurstone's instructions for the 'centroid' method. This involves two departures from the original version. First, the reduced self-correlations (or 'communalities') are estimated by simply taking the highest correlation in each column; and this rough device is applied afresh to the successive tables of residuals. Now a change of this sort is almost bound to upset the sign-relations: the saturations or 'loadings' for the bipolar factors will no longer add up exactly to zero, and hence we cannot expect the symmetry of positive and negative signs to be so neatly preserved. Often the factors fail to appear in the strict order of variance; and occasionally artificial factors may even be introduced to account for the errors produced by the crudeness of the estimates. Secondly, in the original version the routine procedure for 'reflecting signs' (as it is sometimes called) was based on the same principle as is tacitly employed in the method of weighted summation or principal components: in the latter we multiply the coefficients by a set of weights with different signs and different numerical values—the 'best' weights: in the former, we are content with approximate values obtained by treating the weights as numerically equal, but still preserving the original differences in sign. Thurstone's suggestions for 'reflecting signs' appear to be merely a convenient rule-of-thumb improvisation, having no theoretical basis. As a result, they frequently yield a different sign pattern. No doubt, judged from his own special standpoint, the fact that the sign-law¹ may consequently be upset does not really matter; for he regards the distribution of positive and negative saturations as possessing no intrinsic meaning, and never uses them as clues to the appropriate grouping of the tests or traits.

To investigate the presence of the law in actual researches, it is therefore necessary to examine analyses (a) which have been rigorously carried out, (b) which are based on a large number of cases, and (c) which yield a large number of fully significant factors. Unfortunately comparatively few investigations satisfy these conditions.

(1) Where the test-items have been selected to measure only a single psychological characteristic, like the Binet Scale and the Progressive Matrices—both designed to measure intelligence—the available analyses show results conforming reasonably well with the law (8, 11). This is in keeping with Mr. Gabriel's contention, since such tests might be held to obey the requirements of a Simplex structure. So far as the items form a fairly homogeneous series, increasing solely or mainly in complexity, then, as we have seen, the factors will merely be 'difficulty factors'²; and, if the increments in difficulty are approximately uniform throughout the scale, a progressive increase in sign-change is inevitable. In such cases

¹ Curiously enough, if we turn to Thurstone's earliest illustration of his procedure (*Vectors of Mind*, Table 25, p. 117), and re-arrange the tests in the order 10, 2, 5, 9, 6, 4, 3, 1, 8, 7, 14, 17, 15, 11, 18, then the number of sign-changes is (I) 0, (II) 1, (III) 2, (IV) 3, (V) 4, exactly as the law requires. If we re-calculate the saturations by successive approximation, in accordance with the procedure I myself originally proposed, thus ensuring that the bipolars sum to zero, the factor-pattern becomes clearer still: it then corresponds closely to what would seem a very plausible way of classifying the tests (Verbal v. Non-verbal, Verbal Opposites v. Other Verbal, Mathematical v. Non-mathematical, etc.).

² As noted in the article cited, I do not believe that *all* the factors obtained from the Binet Scale are mere difficulty factors: some certainly suggest qualitative difference, and this suggestion is borne out by special studies of the children with large factor-measurements.

I should agree with Mr. Gabriel that the peculiarity reveals little or no fresh information about the psychological nature of mental abilities. We shall all, however, willingly acknowledge the importance of the main point that he wishes to urge, namely, that we have here a new and suggestive device for investigating how far such tests do in fact exhibit the structure intended.

(2) The law seems to be satisfied ¹ in most analyses of intellectual assessments, both for abilities and for attainments : it is clearly discernible in the earliest studies of educational performances and the latest researches with purely cognitive tests (2, 6, 13 and refs.). For emotional characteristics the evidence is more scanty. With the aid of an electronic computer, Dr. Wrigley has made a complete analysis by principal components of one of my own early correlation tables.² His results (given to three decimal places only) are shown in Table IV.

TABLE IV. FACTOR ANALYSIS OF EMOTIONAL ASSESSMENTS
Complete Analysis by Principal Components

Factor	I	II	III	IV	V	VI	VII	VIII
Joy	.779	.189	.295	.135	.476	.139	.087	.018 <i>i</i>
Tenderness	.799	.456	.237	-.128	-.280	.064	-.059	.072 <i>i</i>
Sorrow	.932	.146	-.072	-.112	-.243	.011	.190	-.054 <i>i</i>
Fear	.426	.463	-.685	.359	.061	.001	-.051	-.000 <i>i</i>
Anger	.557	-.686	-.163	.210	-.115	.369	-.000	.024 <i>i</i>
Disgust	.607	-.243	-.395	-.610	.185	-.106	-.014	.024 <i>i</i>
Sociability	.959	-.045	.213	-.038	.032	.026	-.185	-.070 <i>i</i>
Wonder	.730	-.374	.092	.327	-.033	-.460	.015	.024 <i>i</i>
Contribution to variance (per cent.)	4.427	1.151	.853	.700	.417	.383	.084	-.015
Number of sign-changes	0	1	2	4	5	3	4	4

Note.—The small imaginary factor is due to the fact that the assessments for certain traits (Sex and Fear) were based on fewer individuals than the rest : so that the correlations do not all refer to exactly the same sample.

This presents an empirical factor-matrix for exactly eight variables, and so gives a concrete example which can be directly compared with the artificial case illustrated in Table I. In Table IV, it will be seen, the sign-changes do not comply very closely with the law we have been considering. Yet that fact of itself has psychological implications. Factor II, as usual, divides the traits into two contrasted groups—the sthenic (or extraverted) and the asthenic (or introverted) group. But now the two groups are not subdivided *separately*, like those in Table IA. The subdivision is the same for both, namely, into pleasant emotions and

¹ With tests like the Progressive Matrices the arrangement of the items is independently determined, since the order of difficulty or complexity may be inferred either from their construction or from the proportion of children passing them at each age, or even from their saturations for the general factor. With more heterogeneous tests external clues of this sort may be lacking ; and the investigator has greater freedom in arranging the items to fit the sign pattern. Unfortunately with cognitive tests, except in the case of the research on educational tests (6), weighted summation has rarely been used. When simple summation is substituted and when the investigator is at liberty to adopt any order he likes, it is always possible to arrange the first three factors (general and two bipolar) to fit the requirements of the sign-law. Hence, in such cases, the agreement means nothing unless it extends to four or more factors. In a few experiments, however, that have been deliberately planned (e.g., in Moursy's investigation with cognitive tests (13) and some of my own earlier studies), the 'level' of the test in the hierarchy can be determined on antecedent grounds ; and here the evidence is far more convincing.

² The table was one of those given in my paper on 'General and Specific Factors underlying the Primary Emotions,' *Brit. Assoc. Ann. Rep.*, 1915, pp. 694-6. It is reprinted in Table 7.0 by Holzinger and Harman (*Factor Analysis*, 1941, p. 176) ; but, as in my own analysis, the reduced matrix is factorized, and only two factors extracted. I am much indebted to Dr. Wrigley for the more elaborate analysis.

unpleasant ; and consequently, in Table IV, Factor IV no longer repeats the subdivisions found in Factor III, with the last four signs reversed, as happened in Table IB. More recently Dr. Wrigley has factorized in the same way a correlation table for eleven emotional traits (Burt, C., *Journal*, V, p. 199). In this instance the number of sign-changes is 0, 1, 2, 3, 5, 5, 6, 7, 6, 4, 5. The agreement is a little closer ; but the same disturbance of the theoretical pattern is visible.

(3) Finally, it appears that the law is discernible even outside the psychological field. Consider, for example, the results obtained from one of the earliest correlation tables to be factorized, that calculated by Macdonell for seven physical traits measured at Scotland Yard on 3,000 criminals in accordance with the Bertillon scheme. One of the main objects was to work out a practicable filing system for classifying criminals ; and it was this that led Pearson to propose the method of 'principal axes.'¹ Table V gives the results of a complete factorization by the method of simple summation. The factors appear in order of diminishing variance ; and it will be seen that, except for the last factor of all, the number of sign-changes increases regularly throughout the analysis. On the other hand, on analysing the same data by the method of principal components the agreement is not so neat, for reasons I have explained elsewhere (2, p. 109).

TABLE V. FACTOR ANALYSIS OF PHYSICAL MEASUREMENTS
Complete Analysis by Simple Summation

Factor	I	II	III	IV	V	VI	VII
1. Head length	·61	·43	·53	—·39	·00	·00	·00
2. Head breadth	·53	·66	—·25	·20	·32	—·28	·00
3. Face breadth	·66	·51	—·28	·20	—·32	·28	·00
4. Finger	·80	—·41	—·17	—·21	—·19	—·22	—·19
5. Foot	·83	—·36	—·15	—·18	·19	·22	·19
6. Forearm	·82	—·47	·09	·10	—·13	—·18	·19
7. Height	·80	—·36	·23	·30	·13	·18	—·19
Contribution to variance (per cent.)	53·5	22·0	7·7	5·9	4·5	4·4	2·0
Number of sign-changes	0	1	2	3	4	5	(?)

A more recent example, drawn from the same non-psychological field, will be found in the analysis of a set of nine body-measurements obtained from 2,400 recruits for the Royal Air Force. In this case, following the more usual procedure, the analysis was applied to the reduced correlation matrix. Hence a perfect fit is obtainable with six factors only. If the reader examines the factor-matrix, taken just as it stands (9, Table 2, p. 242), it will be seen that the number of sign-changes is 0, 1, 2, 4, 4, 5. The slight discrepancy in factor IV is due to a single trait with a saturation of only —0·04 ; on adjusting the figures in this column by weighted summation, the non-significant negative value is changed to a positive, and the law is then perfectly preserved.

Now the fact that the sign-law appears in so many different fields of work might at first sight suggest that it is after all nothing but a statistical artefact—an inevitable consequence of the procedure employed ; and that would be in keeping with the view, expressed by Thurstone and others, that sign patterns obtained by summational analysis have no concrete

¹ *Biometrika*, I, 1902, pp. 202, 207. This is the earliest instance in which a complete table of inter-correlations is calculated and set out in the now familiar form of a symmetrical determinant or matrix. Pearson himself never actually applied his method to any empirical correlation table, partly because with more than four variables the computations seemed too laborious, and partly because the success of the finger-print method had in the meanwhile superseded the Bertillon scheme. The correlations calculated by Macdonell are reproduced in this *Journal*, II, p. 104, together with my own earlier attempts at analysing the matrix by the method of principal components, simple summation, and other procedures.

meaning for the psychologist. But I believe that such an inference would be entirely mistaken. First of all, it is quite easy to devise classificatory schemes which flagrantly transgress the requirements of hierarchical structure ; and in such cases the sign-law no longer holds good. Hence it is not an inevitable result. Secondly, it is by no means difficult to find empirical instances, like that illustrated in Table V, in which the sign-law is broken by saturations that are nevertheless fully significant statistically. Thus whether it is obeyed or not depends on the nature of the material.

I maintain therefore that, with psychological data, the appearance of a progressive sign-change *does* have important psychological implications. It provides the most convincing way of verifying or refuting the hypothesis that the traits selected can be classified according to a hierarchical scheme. The exceptions are as instructive as the instances of agreement. They indicate where and often why the traits depart from the ideal scheme. The commonest infringements consist in the skipping of some of the factors in which the grouping is identical although the signs are dissimilar ; e.g., a factor like factor III in Table IB may be immediately followed by one like factor V, the analogue of factor IV being skipped. This implies either that the larger group factors are not subdivided separately (as in Table V) or that they are entirely missing and that the smaller group factors (with the general or basic factor) are alone sufficient to account for the correlations observed.

But, as I have tried to show in previous papers, the pattern of signs obtained for the more significant of the summational factors not only supplies an objective basis for classifying those traits that were selected for the analysis, it also provides the safest way of determining the mode of rotation. Accordingly, it is always desirable to re-arrange the tests or traits in the table of bipolar factors so as to bring out their classificatory relations.¹ As a working rule, the computer should seek to re-arrange his list of tests or traits so that the smallest number of sign-changes appears in each column. I do not contend that this is the only appropriate order. In certain cases it may be better to begin each new section or subsection with the positive saturations : otherwise, when similar signs from different sections are brought together (as in factors III and V in Table III above), that may perhaps suggest an erroneous grouping (e.g., in Table IV, that Finger and Face breadth go together).

IV. SUMMARY AND CONCLUSIONS

1. On examining tables of correlations based on psychological data and analysed by factorial methods, it appears that, provided the number of traits, persons, and factors is large enough for the question to be tested, the majority yield factor-matrices conforming fairly closely with the law of progressive sign reversals. This peculiarity, therefore, is not, as has been suggested, restricted to those special types of composite test which exhibit a 'Simplex structure.' Moreover, the same tendency may be discerned in analyses of physical measurements. It thus seems to be characteristic of certain types or sets of measurements derived from the biological or anthropological field, and not merely from the psychological.

2. The law is not, as certain writers have contended, an automatic consequence of the factorial procedure employed, nor is it a mere mathematical phenomenon possessing no psychological implications. It is a logical consequence of the structural relations obtaining among the factors which underlie the material ; it appears whenever the factors represent hypothetical tendencies that can be classified and subclassified as parts of a fairly complete hierarchical scheme. When the traits measured are not classifiable in this way, the factor-matrix no longer obeys the law.

3. In this country the earliest applications of summational methods of analysis were planned and carried out by investigators who held, on *a priori* grounds, that many of the more fundamental psychological and biological characteristics do in fact form part of a hierarchical classification ; and the methods were to a large extent devised and applied in order to secure empirical evidence for testing this hypothesis. The sign pattern exhibited

¹ A good instance is provided by the twenty-one tests discussed in a previous paper in this *Journal* (IV, p. 21, Table II) which two investigators have interpreted in different ways (cf. this *Journal*, V, p. 12). As I ventured to indicate at the time, as soon as the tests are re-arranged according to the sign pattern, the main group factors become clear at once (*ibid*, p. 15, and Table VI, p. 18).

by the summational or 'bipolar' factor-matrix has also been used as the objective basis for classifying the traits actually tested; and the result has been to show (a) what traits are classifiable in this way, and (b) what is the detailed nature of the classification so revealed. Hence, unless there are overriding reasons to the contrary, the bipolar factor-matrix should always be arranged so as to minimize the number of sign-changes in each column, and thus demonstrate how far the general rule is actually obeyed.

REFERENCES

1. Macdonell, W. R. (1902). 'On criminal anthropometry and the identification of criminals.' *Biometrika*, I, 177-227.
2. Burt, C. (1911). 'Experimental tests of higher mental processes and their relation to general intelligence.' *J. Exp. Ped.*, I, 93-112.
3. Spearman, C., and Hart, B. (1912). 'General ability: its existence and nature.' *Brit. J. Psychol.*, V, 51-84.
4. Thurstone, L. L. (1935). *The Vectors of Mind*. Chicago: University Press.
5. Burt, C. (1938). 'The unit hierarchy and its properties.' *Psychometrika*, III, 151-68.
6. Burt, C. (1939). 'The relations of educational abilities.' *Brit. J. Educ. Psychol.*, IX, 45-71.
7. Burt, C. (1940). *The Factors of the Mind*. London: University of London Press.
8. Burt, C., and John, E. (1942). 'A factorial analysis of the Binet tests.' *Brit. J. Educ. Psychol.*, XII, 117-21, 156-61.
9. Banks, C., and Burt, C. (1947). 'A factor analysis of body measurements for British adult males.' *Ann. Eugen.*, XIII, 238-56.
10. Burt, C. (1949). 'Alternative methods of factor analysis and their relations to Pearson's method of principal axes.' *Brit. J. Psychol., Stat. Sect.*, II, 98-121.
11. Keir, G. (1949). 'Progressive matrices applied to school children.' *Brit. J. Psychol., Stat. Sect.*, II, 140-50.
12. Vernon, P. E. (1950). *The Structure of Human Abilities*. Methuen.
13. Moursy, E. M. (1952). 'The hierarchical organization of cognitive levels.' *Brit. J. Psychol., Stat. Sect.*, V, 151-80.
14. Burt, C. (1952). 'Bipolar factors as a cause of cyclic overlap.' *Brit. J. Psychol., Stat. Sect.*, V, 198-202.
15. Burt, C. (1953). 'Scale analysis and factor analysis.' *Brit. J. Stat. Psychol.*, VI, 5-24.
16. Gabriel, K. R. (1954). 'The Simplex structure of the progressive matrices test.' *Brit. J. Stat. Psychol.*, VII.

A COMPARISON OF TWO METHODS OF FACTORIZING TEST DATA

By DUNCAN HOWIE

New England University College, University of Sydney

I. Problem. II. Illustrative Example. III. Comparison of Burt and Thurstone Factors.
IV. Summary and Conclusions.

I. PROBLEM

Earlier Comparisons of the Burt and Thurstone Techniques. In psychology, as in other sciences, when two or more investigators, analysing the same material, draw opposite inferences, the commonest explanation is that the difference in their conclusions results from a difference in their methods of analysis. This has been one of the chief sources of controversy in factorial researches from the very start. Thus, Burt, on analysing correlation tables from his own data and from Simpson's monograph, claimed to find group factors as well as a general factor; while Spearman, re-analysing the same tables with his two-factor method, could find only a general factor (9). Conversely, Thurstone, re-analysing a correlation table in which one of Burt's students had discovered a large general factor, found that there is nothing but a simple structure of group factors (11). To-day probably all psychologists would admit the existence of group factors, though they disagree over their nature. On the other hand, most of those who adopt Thurstone's techniques appear to find no grounds for assuming a general factor as well.

In his recent book (5) Professor Vernon devotes a long Appendix to the question of 'General-plus-Group Factors *versus* Multiple Factor Theories,' and offers seven reasons "why general-plus-group factor solutions are considered superior." His arguments are based chiefly on abstract considerations. Here it is proposed to attempt a concrete comparison of the actual figures obtained by applying *both* methods of analysis to the *same* data. In the earlier discussions like those just cited, the discrepancies arose out of the analysis of material obtained with comparatively small batteries of tests; and it has often been maintained that, if the batteries were enlarged, many of the differences would be automatically settled. However, after publishing a research (6) in which as many as 36 variables were employed, the present writer was confronted with very similar criticisms. And it therefore seems desirable to enquire how far a change in factorial procedure would produce a change in the conclusions then drawn.

Only two other comparisons along these lines have, it would appear, been carried out. The large correlation table, factorized by Thurstone to obtain a 'simple structure' of 'primary abilities' (10), was re-factorized by Eysenck with Burt's group factor technique, and found to contain a general factor as well as a small number of group factors corresponding with the Thurstone 'primary abilities' that have fullest statistical significance (4). More recently two French investigators have applied *le méthode de Burt* and *l'analyse multifactorielle de Thurstone* to four correlation tables derived from testing nearly 2,000 boys and girls; they conclude that, provided it is pressed to its logical outcome (the extraction of second order factors from the correlated primary), Thurstone's procedure leads "dans le cas de nos données, et vraisemblablement dans beaucoup d'autres cas, aux solutions orthogonales en *facteur général* et *plusieurs facteurs de groupe*, telles que Burt les préconise de longue date" (8; a summary will be found in this *Journal*, VI, 1953, pp. 126-8). Since in nearly all such cases the general factor actually accounts for more of the total variance than all the group factors put together, it is, they contend, as illogical to reject the assumption of a general factor (with Thurstone's followers) as it is to reject the assumption of group factors (with Spearman's followers).

This paper will be concerned primarily with two alternative methods of orthogonal analysis. The merits and demerits of the later extensions suggested by Thurstone—introducing oblique in preference to orthogonal factors and applying a further analysis to secure second order factors—will not concern us. They have been discussed in some detail by previous writers ; and here it seems simplest to deal with the more fundamental differences first.

II. ILLUSTRATIVE EXAMPLE

Data. My original enquiry was concerned with the relation between associative fluency and concentration, and their influence on scholastic ability. 34 tests were applied to 252 school children (119 boys, 133 girls, mean age 12 years 8 months, standard deviation 1 year 1·5 months). Two main conclusions appeared to emerge from the analysis : (1) no support was found for the hypothesis that the factor of ‘associational fluency’ either facilitates or hinders scholastic performance ; (2) strong support was obtained for the hypothesis that what is popularly described as ability to concentrate or at any rate its core, is represented by a factor that may broadly be described as a factor of ‘reasoning’—shown in tests “which require the organization of ideational material under conditions involving the maintenance of a directive set.” The possibility arises that the interpretations and conclusions offered were an outcome of the particular factor technique used, namely, Burt’s method of group factor analysis. To meet this objection the writer has since re-analysed the initial data by means of Thurstone’s technique. For further details as to population, tests, correlations, and fuller discussion of the literature of the special problem, the reader is referred to the earlier paper (7).

Methods. (i) *The Burt Procedure.* The method followed was that described by Burt in *The Factors of the Mind* (Appendix on ‘Working Methods,’ pp. 277f.). Two minor modifications were introduced. (a) The initial groupings¹ were made by the use of Holzinger’s B coefficients. (b) After extracting the basic factor, certain tests, whose residuals showed significant linkage with an additional group, were included in that group for purposes of computation. The pattern is thus one of group factors with limited overlap. With 36 variables, the degree of inflation in the basic factor saturations, and consequent depreciations in the group factor saturations where procedure (b) was used, would be slight.

(ii) *The Thurstone Procedure.* The centroid axes were calculated and then rotated orthogonally in pairs, until the negative values had been reduced to a minimum and the number of zeros increased to a maximum. The resulting pattern gave a satisfactory approach to the requirements of a ‘simple structure’ as defined by Thurstone (10).

In each analysis, after eight factors had been extracted, the mean of the residuals was zero and the standard deviation less than the standard error of a zero tetrachoric correlation.

III. COMPARISON OF BURT AND THURSTONE FACTORS

Results. In order to illustrate in the concrete the various consequences of the different factorial approaches, the results of the two analyses are presented in parallel in Table I. It will be observed that with the Thurstone analysis the variances of the group factors are larger ; this is due partly to the fact that, as noted above, Burt’s procedure was not extended to include his method of obtaining minor saturations for all the variables, but mainly to the fact that Thurstone’s procedure yields no large general factor. A few detailed comments on the several factors are necessary to bring out the main points of agreement and disagreement.

Factor I. The first factor resulting from the Burt analysis is clearly a general factor of the usual type. By virtue of the requirements of ‘simple structure,’ it does not appear in the Thurstone analysis. Instead the partial correlations resulting from Burt’s general factor are accounted for by the wide overlapping of the Thurstone group factors, especially

¹ At the time when the investigation was carried out, Burt’s method of computing overlapping group factors by arithmetical rotation was not available ; nor was the writer then aware of the importance of grouping according to the lines of division indicated by the bipolar analysis.

factors II and IX. Factor IX, which appears solely in the Thurstone analysis, has fairly large saturations for 27 of the variables; and, where this factor shows negative or near zero correlations, Thurstone's factor II has fairly large saturations.

Factor II. This factor was interpreted as a verbal factor identifiable with that described by Burt (1, pp. 184f.). It appears to be primarily involved in the mechanics of the reading-process required by many of the tests. The Burt saturations are all consistent with this interpretation. It will be seen that the Thurstone analysis agrees with the Burt in showing substantial saturations for Reading (Word Knowledge), Reading (Detail), Reading (Speed), and Listing; but a noticeable discrepancy occurs in the absence of a significant saturation for Spelling in Composition. The Thurstone saturations for Substitution, Memory Digits Backward, Memory Digits Forward, and possibly Concentration, might, in line with the above interpretation, be taken as indicating that the particular aspect of the mechanics of the reading-process involved in this factor may be roughly identified with 'reading span.' As against this, however, the saturations for Classification, Free Chain Association, Fluency—Topics and Composition Assessment, suggest a broader interpretation on the lines of Thurstone's observation: "it is quite likely that the factor V (i.e., Verbal) will be determined largely in terms of verbal manipulation of ideas as they occur in sustained verbal discourse." (11). The Thurstone factor also includes tests where speed of writing would be an influence in the score; hence I find it difficult to explain the saturation for Test 17 (Writing Speed). The large saturation for the Non-Verbal Test appears quite inexplicable.

It is clear, then, that, despite agreements in the two descriptions, the Thurstone factor is more complex than the Burt factor.

Factor III. Here the two analyses agree in showing saturations for Problem Arithmetic, Mechanical Arithmetic, Listing, Classification, and Reading (Word Knowledge). The common feature of these tests is that they involve discrimination and analysis of data in order to make an appropriate selection or to effect an appropriate organization in terms of some governing principle (*Aufgabe*). The further saturations in the Thurstone analyses—those for Reading Detail, Reading Speed, Jumbled Paragraphs, Memory-Prose Passage, and Word Building—are consistent with this characteristic. Like the Burt factor the Thurstone factor has negligible, or even negative, saturations for the fluency tests. Thus neither factor enters into those tests which require associative material to be rapidly produced and which impose no condition of relevancy on the associations. In general terms, therefore, factor III may be regarded as a reasoning factor;¹ but its rational characteristic seems to relate, not to a capacity for logical inference (e.g., induction or deduction), but to an ability to order material in terms of some governing principle or some constraining form. It will be noted that, with Burt's technique, this factor has a large negative saturation (—51) for Sex; with Thurstone's the saturation is also negative, but barely significant (—14).

Factor IV. This factor shows a closer agreement between the two versions than any of the others. The spread of Thurstone saturations beyond the Burt is limited to a saturation for Fluency Predictions. For the tests taken from Cattell—Fluency-Completing Forms, Fluency-Word Series, Fluency-Pictures, Fluency-Ink Blots—the saturations are high in both analyses. The saturations for a test introduced by the writer (Matches Construction) are in keeping with this. These figures clearly define the factor as involving associational facility, and support Cattell's claim that the essential characteristic of the so-called 'fluency' factor is a fluency of associations under restrictions (3, p. 427, 8, p. 258). The saturations for Writing Speed are probably due to the fact that for children of this age speed of writing helps to achieve a good score in these tests; it is, however, possible that they indicate a relation between fluency and a more general capacity of speed (1, pp. 181, 195 and refs.; 3, pp. 420 ff.).

¹ Burt has pointed out (1, p. 197) that the 'capacity for cognizing a number of items as an organized whole' appears very similar to what teachers call 'attention' and the faculty psychologists 'concentration.' In the earlier paper (6) the writer proposed the term 'conceptual concentration' to describe the factor mentioned above: such a phrase, however, might be misunderstood to imply a temperamental characteristic; and this could hardly be supported from the evidence available from his data. For the same reason the description of factor V as 'perceptual concentration' is now altered.

TABLE I. SATURATIONS FOR BURT'S GROUP FACTORS AND THURSTONE'S ROTATED FACTORS

Tests	Factors												Communality						
	I	II		III		IV		V		VI		VII		VIII		IX			
	B	B	TH	B	TH	B	TH	B	TH	B	TH	B	TH	B	TH	B	TH		
1. Reading (Word Knowledge)	.67	.65	.46	.32	.59		-.07	.16		.19		.28		-.16		.33		.97	.84
2. Reading (Detail)	.59	.57	.51	.20	.53		-.04	.10		.15		.17		-.18		.32		.72	.74
3. Reading (Speed)	.59	.52	.62	.39	.39		-.02	.06		.01		.17		-.15		.32		.62	.69
4. Spelling (in Composition)	.44	.52	.08	.18	.25		-.08	.12		.29		.62		-.23		.29		.34	.38
5. Grammar (in Composition)	.32	.26	.01	.02	.15		-.02	-.09		.22		.53		-.18		.14		.45	.45
6. Listing	.53	.61	.36	.42	.36		-.17	.20		.13		.12		-.09		.09		.62	.63
7. Jumbled Paragraphs	.66	.41	.04	.67	.61		-.04	.09		.22		.12		.11		.54		.44	.56
8. Problem Arithmetic	.48	.32	.02	.53	.51		-.16	.24		.17		.12		.14		.14		.43	.47
9. Mechanical Arithmetic	.50	.44	.40	.40	.45		-.05	.13		.13		.08		.25		.32		.42	.61
10. Classification	.44	.31	.34	.34	.35		-.05	.05		.12		.34		.21		.19		.31	.43
11. Non-Verbal Test	.10	.22	.22	.30	.24		-.58	-.08		-.16		-.13		.18	.34	.26		.82	.58
12. Fluency (Completing Forms)	.43	.22	.16	.15	.15		.56	.50		.02		.02		.20		.39		.50	.54
13. Fluency (Word Series)	.33	.16	.34	.19	.19		.50	.69		.19		.06		-.09		.02		.42	.65
14. Fluency (Pictures)	.38	.34	.05	.05	.23		.31	.23		.25		.04		-.04		.47		.28	.40
15. Fluency (Topics)	.23	.12	-.02	.02	.02		.53	.51		-.25		.20		.00		.32		.31	.48
16. Fluency (Ink Blots)	.54	.40	.12	-.02	.31		.41	.29		.42		.16		-.20		.00		.47	.62
17. Writing Speed	.21	.21	.14	.14	.23		.48	.25		.27		.04		.28		.20		.21	.46
18. Matches Construction	.31	.06	.08	.17	.19		.32	.10		.17		.05		.17		.59		.44	.66
19. Sex (Feminine)	.64	.35	.35	.25	.25		.10	.70		.32		.42		.05		.26		.47	.50
20. Composition (No. of Words)	.58	.45	.45	.00	.23		.23	.62		.30		.23		-.02		.46		.83	.77
21. Concentration	.55	.29	.29	.17	.17		.19	.56		.47		.08		.13		.17		.62	.71
22. Free Chain Association	.05	.29	.43	.17	.17		.03	.43		.26		.17		.18		.00		.32	.54
23. Age	.59	.19	.19	.08	.08		-.14	.40		.37		.20		.07		.36		.53	.61
24. Substitution	.25	.25	.06	.16	.16		-.06	.10		.05		.24		.04		.39		.22	.51
25. Number Checking	.51	.43	.39	.14	.14		.10	.15		.49		.18		.07		.10		.44	.54
26. Memory (Movements Seqs.)	.43	.27	.34	.05	.23		-.03	.23		.30		.17		.02		.35		.34	.44
27. Memory (Digits Backward)	.27	.26	.26	.43	.43		.04	.09		.26		.33		.22		.16		.57	.57
28. Memory (Digits Forward)	.52	.26	.26	.28	.28		.11	.08		.22		.33		.25		.35		.34	.57
29. Memory (Prose)	.07	.07	.07	.27	.27		.03	.14		.08		.14		.07		.16		.32	.22
30. Memory (Nonsense Sylls.)	.63	.26	.26	.27	.27		.14	.01		.03		.62		.27		.24		.66	.68
31. Composition (Assessment)	.49	.20	.20	.11	.11		.14	.18		.11		.45		.36		.24		.44	.55
32. Composition (Neatness)	.41	.24	.24	.20	.20		.03	.18		.03		.30		.22		.04		.20	.28
33. Composition (Length of Sentences)	.60	.19	.19	.20	.20		.30	.27		.23		.35		.38	.65	.29		.78	.64
34. Fluency (Predictions)	.49	.16	.16	.08	.08		.01	.10		.01		.20		.46		.40		.45	.42
35. Memory (Pictures)	.45	.02	.02	.32	.32		-.05	.15		.14		.01		.19		.25		.24	.36
36. Word Building																			
Percentage of variance	22.0	4.3	9.2	4.5	7.9	5.1	6.9	4.6	5.8	2.9	5.8	2.4	7.3	2.4	4.2	8.5		48.3	55.7

B = Burt Group Factor. TH = Thurstone Rotated Factor.

Factor V. With this factor the two analyses closely agree in their saturations for Concentration, Substitution, Number Checking, Free Chain Association, Speed of Writing, and Age. The first three suggest a relation with Thurstone's factor of perceptual speed, viz., "a facility in perceiving detail that is embedded in irrelevant material" (10, p. 81), "readiness to discover and identify perceptual material" (11, p. 2). The saturations for Free Chain Association and Speed of Writing, and the Thurstone saturation for Number of Words in Composition, appear to indicate that the factor is rather a speed factor in habitual or routine operations; and the saturations for Age perhaps imply that the older the child, the more the test materials tend to be taken as matters of practice and habit. Like the fluency factor, this factor shows little spread of the Thurstone saturations beyond the Burt. With the exception of the slight and doubtful values for the two Arithmetic tests, the Thurstone analysis supports the conclusion already drawn from the Burt analysis: namely, that this factor has a very restricted function and does not enter into the description of the scholastic abilities tested.

Factor VI. Factor VI appears to be identifiable with rote memory as tested by immediate recall (cf. 1, p. 177 and refs.; 8, p. 236). This is shown by the fact that in the Burt analysis it has substantial saturations for all five memory tests and for one or two scholastic abilities into which rote memory might be expected to enter (Spelling and Grammar). The further Thurstone saturations—for Concentration, Substitution, and Number Checking—do not seem inconsistent with this interpretation. In neither analysis, however, does this factor greatly affect scores for the more important scholastic tests.

Factor VII. In both analyses the higher saturations for this factor are restricted to marks for certain features of the Composition test, and suggest that the assessments may have been appreciably influenced by a complex of impressions derived from the more mechanical aspects of the children's compositions. This result naturally suggests a comparison with the more complex type of verbal factor reported by previous investigators: e.g., in an early study of *The Distribution and Relations of Educational Abilities* (1917, p. 59), Burt reported, in addition to a more elementary type of verbal factor (namely, what he called a 'linguistic' factor, facilitating reading and spelling), a second factor having specially high saturations in tests involving written composition (cf. 1, p. 185). Factor VII is probably related to this factor; but the present data do not justify any more precise identification.

Factor VIII. The saturations for this factor suggest that it may be identified with that type of creative fluency which Burt has called productive imagination (as distinct from reproductive) (1, p. 180). Cattell describes it as "fluency of association with minimal direction," and a similar factor is reported by Carroll (2) and Gibb (6). It apparently implies that, under conditions of minimal direction, the associations will be governed by an organizational set within the individual: he will, as it were, structure his associational field. The highest saturation is for Test 35 (Memory for Pictures): in this a picture showing 42 different objects was presented for thirty seconds; and doubtless the child's success is largely due to his organizational memory. The Thurstone saturation for Problem Arithmetic might perhaps be explained in the same way; but the high Thurstone saturation for Neatness in Composition appears quite anomalous.

There are thus seven group factors representing 'special aptitudes' or 'primary abilities' which appear with both types of analysis. Most of them might be identified with the group factors, described under slightly more general titles, in the list given in Burt's *Measurement of Mental Capacities* (1925), and all are to be found among the group factors mentioned in his later review (1). In the past such factors have usually been detected in separate researches; here they appear in a single large battery of tests, and thus, it is suggested, can be described and discriminated with much greater detail and precision.

Factor IX. While factor I is found only with Burt's group factor analysis, so factor IX is found only in the 'simple structure' produced by the Thurstonian rotation. Its highest saturation is for Sex; and it accounts for 53 per cent. of the common factor variance of this variable. The tests with significant saturations are for the most part those in which the girls did better than the boys. On looking at the table the reader will see that these tests are in keeping with what is known about sex differences from other investigations. In the literature of the subject a sex factor as such seldom appears, since the sexes are usually separated when correlation tables are calculated, and the evidence for such differences is commonly based on external rather than internal criteria.

IV. SUMMARY AND CONCLUSIONS

1. A series of some 36 variables (chiefly cognitive and scholastic) was factorized by Burt's group factor method. To meet the objection that the factors found and the inferences drawn might possibly be the effects of the factorial technique adopted, the data have now been re-factorized by Thurstone's procedure. It is found that in the main the same factors emerge. Thus the factors that appear in the 'simple structure' seem to agree in their description with those previously found with the earlier procedure, though the Thurstone procedure gives rise to certain discrepancies and difficulties.

2. The factors revealed by both analyses appear to be the following : (i) a verbal factor ; (ii) a reasoning factor (ability to order material under the control of some formal principle) ; (iii) a factor of fluency (facility of association with a minimum of control) ; (iv) a speed factor in routine operations ; (v) rote memory ; (vi) mechanical aspects of composition ; (vii) creative imagination.

3. The most conspicuous difference is that Burt's method reveals a general factor, identifiable with 'intelligence' and accounting for 22 per cent. of the variance, while Thurstone's method reveals a sex factor accounting for 8.5 per cent. of the variance.

4. The Thurstone analysis fully supports the main conclusions derived from the Burt analysis : (i) the factor of fluency is apparently not involved in the scholastic abilities ; (ii) the only indication that a factor of concentration enters into the scholastic abilities is the appearance of what might perhaps better be described as a reasoning factor, i.e., a factor representing ability to order ideational material in terms of some constraining form.

5. The comparison of the two analyses raises a fundamental issue : how is one to choose between, or evaluate, different factorial techniques in terms of the descriptions of the data they afford ? On grounds of psychological intelligibility there seems little preference for either one or the other. The Thurstone analysis throws up some anomalous saturations. On the other hand, it does on occasion fill out the factorial picture, as in the case of factor III and factor VI. However, by adopting Burt's method of dealing with overlapping factors, this could probably be achieved with his procedure. Because of a certain neatness and parsimony, the writer would incline to the Burt description ; but this neatness may possibly be obtained at the expense of a loss of descriptive completeness, as in the sex difference factor here revealed by the Thurstone analysis. Moreover, the Burt procedure allows for a general factor which the Thurstone procedure spreads over the remaining factors. From an eclectic standpoint it might perhaps be safer in the present state of our knowledge to confine the description of factors to those tests in regard to which the different techniques show agreement ; but this should not mean avoiding the challenge furnished by the divergences, since it is often at the points of divergence that progress in research occurs.

REFERENCES

1. Burt, C. (1949). 'The structure of the mind : a review of the results of factor analysis.' *Brit. J. Educ. Psychol.*, XIX, 101-11, 176-99.
2. Carroll, J. B. (1941). 'A factor analysis of verbal abilities.' *Psychometrika*, VI, 279-308.
3. Cattell, R. B. (1946). *Description and Measurement of Personality*. Harrap.
4. Eysenck, H. J. (1939). 'Critical notice of Thurstone's *Primary Mental Abilities*.' *Brit. J. Educ. Psychol.*, IX, 270-75.
5. Vernon, P. E. (1950). *The Structure of Human Abilities*. Methuen.
6. Gibb, C. A. (1942). 'Personality traits by factor-analysis.' *Aust. Journ. Psychol. Philos.*, XX, 86-110.
7. Howie, D. (1950). 'Scholastic aptitude, fluency and concentration,' *Aust. J. Psychol.*, II, 100-13.
8. Reuchlin, M., and Valin, E. (1953). *Tests collectifs du Centre de Recherches*. Paris : Institut National d'Orientation Professionnelle.
9. Spearman, C. (1927). *Abilities of Man*. Methuen.
10. Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago : University Press.
11. Thurstone, L. L. (1938). 'The perceptual factor.' *Psychometrika*, III, 1-17.
12. Thurstone, L. L. (1946). 'The factor analysis of body types.' *Psychometrika*, XI, 15-21.

THE CORRELATION BETWEEN VARIATE-VALUES AND RANKS IN SAMPLES FROM A CONTINUOUS DISTRIBUTION

By ALAN STUART

Division of Research Techniques, London School of Economics

I. *Introduction.* II. *The Correlation between Variate-Values and Ranks.* III. *Bounds for C.* IV. *Particular Distributions.* V. *The Sample Correlation between Variate-Values and Ranks.* VI. *The Asymptotic Relative Efficiencies of Tests against Normal Alternatives.*

SUMMARY

Methods involving rankings have been used by psychologists for half a century ; and, although in their early development they were subjected to severe criticism, recent years have brought their rehabilitation and extension in new directions. Current interest in ranking methods has been displayed in several papers published in this *Journal* and very recently in the correspondence by Burt and Morris (1953). The present paper investigates theoretically the magnitude of the correlation between observations and the rank values by which they are sometimes replaced. It is found that in samples from the two most commonly considered types of statistical distribution, the normal and the uniform, this correlation is very high. With a sample as small as 25, it is .94 for the normal distribution and .96 for the uniform distribution ; and these values increase, with sample size, towards limits of .98 and 1 respectively. In these circumstances there is justification for replacing the original observations by their ranks, since the consequent saving in computation entails very little loss of efficiency. The results of the investigation also turn out to have bearing on other matters of theoretical interest.

I. INTRODUCTION

If N samples of size n are drawn from a continuous distribution with finite variance and the sample values are ranked from 1 to n in each sample, the correlation between variate-values and ranks may be calculated for the complete set of nN observations. Its limiting value, as the number of samples N tends to infinity, is called the correlation between variate-values and ranks in samples of n , and denoted by C_n . The limiting value of C_n as the sample size n tends to infinity, which is denoted by C , is a constant multiple of the ratio of the mean difference of the distribution to its standard deviation, and is therefore non-negative, while for any continuous distribution $C_n = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}} C$ exactly.

C takes the value 1 for the uniform distribution, and $(3/\pi)^{\frac{1}{2}}$ for the normal distribution, but it may be arbitrarily small for particular distributions, and the gamma distribution provides an example when its parameter tends to zero. A lower bound for C is obtained for bounded frequency functions $f(x)$, and from this it follows that a necessary (but not sufficient) condition that C be zero is that $f(x)$ be unbounded at some point within its range.

As a consequence of this work, a lower bound is obtained for the largest ordinate of any bounded continuous $f(x)$ in terms of its standard deviation.

Finally, the results are used to unify the derivations of values for the asymptotic relative efficiencies, against normal alternatives, of several distribution-free test statistics which are essentially efficient estimators with variate-values replaced by ranks.

II. THE CORRELATION BETWEEN VARIATE-VALUES AND RANKS

N samples of size n are drawn from a continuous population with mean μ and variance σ^2 . In each sample, the observations are ordered, the i 'th smallest value, denoted by $x_{(i)}$, being given the rank i . For the complete set of nN observations, we may calculate the covariance between the variate-values and their corresponding ranks (denoted by μ_{11}), the variance of the variate-values (denoted by μ_{20}) and the variance of the ranks (denoted by μ_{02}).

Thus :

$$\begin{aligned}\mu_{11} &= \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N \left\{ i - \frac{1}{2}(n+1) \right\} \left\{ x_{(i)j} - \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N x_{(i)j} \right\} \\ &= \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N i x_{(i)j} - \frac{1}{2}(n+1) \cdot \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N x_{(i)j} .\end{aligned} \quad \dots (1)$$

and

$$\mu_{20} = \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N \left\{ x_{(i)j} - \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N x_{(i)j} \right\}^2 = \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N x_{(i)j}^2 - \left\{ \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N x_{(i)j} \right\}^2 \quad (2)$$

while

$$\mu_{02} = \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N \left\{ i - \frac{1}{2}(n+1) \right\}^2 = \frac{1}{12}(n^2 - 1) . \quad \dots (3)$$

When the number of samples, N , tends to infinity, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N x_{(i)j}^r = E \left\{ x_{(i)}^r \right\} , \quad (r = 1, 2)$$

so that, from (1) and (2),

$$\lim_{N \rightarrow \infty} \mu_{11} = \frac{1}{n} \sum_{i=1}^n i E \left\{ x_{(i)} \right\} - \frac{1}{2}(n+1) \cdot \frac{1}{n} \sum_{i=1}^n E \left\{ x_{(i)} \right\} , \quad \dots (4)$$

and

$$\lim_{N \rightarrow \infty} \mu_{20} = \frac{1}{n} \sum_{i=1}^n E \left\{ x_{(i)}^2 \right\} - \left[\frac{1}{n} \sum_{i=1}^n E \left\{ x_{(i)} \right\} \right]^2 . \quad \dots (5)$$

Now the distribution of $x_{(i)}$, say $G(x)$, is seen combinatorially to be

$$dG(x) = n \binom{n-1}{i-1} \left\{ F(x) \right\}^{i-1} \left\{ 1 - F(x) \right\}^{n-i} dF(x) ,$$

where $F(x)$ is the distribution function of the unordered observations.

Thus

$$\frac{1}{n} \sum_{i=1}^n E \left\{ x_{(i)}^r \right\} = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} x^r dG(x) , \quad (r = 1, 2)$$

and, provided that the interchange of summation and integration operations is permissible, this is

$$= \int_{-\infty}^{\infty} \left\{ \sum_{i=1}^n \binom{n-1}{i-1} F^{i-1} (1-F)^{n-i} \right\} x^r dF ,$$

and the expression in braces under the integral sign is simply the sum of the probabilities in a binomial distribution with parameters F and $(n-1)$, and therefore equals 1, giving

$$\frac{1}{n} \sum_{i=1}^n E \left\{ x_{(i)}^r \right\} = \int_{-\infty}^{\infty} x^r dF = \begin{cases} \mu & (r=1) \\ \sigma^2 + \mu^2 & (r=2) \end{cases} \quad \dots (6)$$

Similarly, we find

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n i E \left\{ x_{(i)} \right\} &= \int_{-\infty}^{\infty} \left\{ (n-1)F + 1 \right\} x dF \\ &= (n-1) \int_{-\infty}^{\infty} x F dF + \mu, \quad \dots (7) \end{aligned}$$

This last integral converges whenever σ^2 is finite.

Putting (6) and (7) into (4) and (5), we obtain

$$\left. \begin{aligned} \lim_{N \rightarrow \infty} \mu_{11} &= (n-1) \left\{ \int_{-\infty}^{\infty} x F(x) dF(x) - \frac{1}{2} \mu \right\} \\ \lim_{N \rightarrow \infty} \mu_{20} &= \sigma^2. \end{aligned} \right\} \quad (8)$$

We now define the correlation between variate-values and their corresponding ranks in samples of size n as

$$C_n = \lim_{N \rightarrow \infty} \frac{\mu_{11}}{(\mu_{20} \mu_{02})^{\frac{1}{2}}} = \frac{\lim_{N \rightarrow \infty} \mu_{11}}{(\lim_{N \rightarrow \infty} \mu_{20})^{\frac{1}{2}}}; \quad \dots (9)$$

and using (3) and (8), (9) becomes

$$C_n = \left\{ \frac{12(n-1)}{\sigma^2(n+1)} \right\}^{\frac{1}{2}} \left\{ \int_{-\infty}^{\infty} x F(x) dF(x) - \frac{1}{2} \mu \right\}. \quad \dots (10)$$

If, now, we also allow the sample size n to tend to infinity, we obtain

$$C = \lim_{n \rightarrow \infty} C_n = \left(\frac{12}{\sigma^2} \right)^{\frac{1}{2}} \left\{ \int_{-\infty}^{\infty} x F(x) dF(x) - \frac{1}{2} \mu \right\}. \quad \dots (11)$$

so that

$$\frac{C_n}{C} = \left(\frac{n-1}{n+1} \right)^{\frac{1}{2}}. \quad \dots (12)$$

Since C_n always approaches its limit C very quickly, we shall henceforth consider only C , whose evaluation in any particular case is a matter of performing the integration in (11).

III. BOUNDS FOR C

If we denote the expression in braces in (11) by $\Delta/4$, it may be verified that

$$\Delta = 4 \int_{-\infty}^{\infty} x \left\{ F(x) - \frac{1}{2} \right\} dF(x), \quad \dots (13)$$

$$= 2 \int_{-\infty}^{\infty} F(x) \left\{ 1 - F(x) \right\} dx, \quad \dots (14)$$

$$= 2 \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^x (x-y) dF(y) \right\} dF(x) . \quad . . . (15)$$

(15) defines Gini's coefficient of mean difference Δ , which is seen from (14) or (15) to be non-negative. Thus, from (11) and (13),

$$C = 3^{\frac{1}{2}} \Delta / (2\sigma) ; \quad . . . (16)$$

and since C is a correlation coefficient, and Δ is non-negative,

$$0 \leq C \leq 1 , \quad . . . (17)$$

in agreement with intuition.*

Using (13), Nair (1936) gave the values of Δ for the uniform, normal, and exponential distributions. Since no derivation is given in his paper, the results are set out in section 4 below, where the exponential distribution becomes the special case of the gamma distribution with parameter 1.

(14) enables us to set a lower bound for the value of C if the continuous frequency function $f(x)$ is everywhere finite. Let $\sup\{f(x)\}$ be the largest value attained by $f(x)$. Then

$$\begin{aligned} \Delta &= 2 \int_{-\infty}^{\infty} F(x) \{1 - F(x)\} dx = 2 \int_{-\infty}^{\infty} F(x) \{1 - F(x)\} \frac{dF(x)}{f(x)} \\ &\geq \frac{2}{\sup\{f(x)\}} \cdot \int_0^1 F(1-F) dF \\ &\geq \frac{1}{3 \sup\{f(x)\}} . \quad . . . (20) \end{aligned}$$

Thus, from (16) and (20),

$$C \geq [2 \cdot 3^{\frac{1}{2}} \sigma \sup\{f(x)\}]^{-1} . \quad . . . (21)$$

For the normal distribution, e.g., (21) gives $C \geq .72$, against the exact value of .98 derived in section IV.

The product $\sigma \sup\{f(x)\}$ is, of course, scale-free, and (21) sets a lower bound for C in terms of $f(x)$ which is certainly non-zero. Thus a necessary condition for C to be zero is that $f(x)$ be unbounded somewhere within its range, though we shall see in section IV that this is by no means a sufficient condition.

As a consequence of (17) and (21), we have the inequality

$$\sigma \sup\{f(x)\} \geq 1/(2 \cdot 3)^{\frac{1}{2}} , \quad . . . (22)$$

which is clearly an equality in the case of the uniform distribution, where $f(x)$ is a constant. (22) provides a lower bound for the largest ordinate in terms of the standard deviation.

* (16) and (17) imply that

$$\Delta \leq 2\sigma/3^{\frac{1}{2}} , \quad . . . (18)$$

which is an improvement on the upper bound for Δ obtainable directly from the Cauchy-Schwarz inequality, namely

$$\Delta \leq 2^{\frac{1}{2}} \sigma . \quad . . . (19)$$

However, our proof of (18) holds for continuous distributions only, while (19) holds for any distribution.

IV. PARTICULAR DISTRIBUTIONS

(a) *Uniform distribution.* The frequency function

$$dF(x) = dx/k \quad (0 \leq x \leq k)$$

$$\text{has } \sigma^2 = \frac{k^2}{12};$$

while, from (15),

$$\Delta = (2/k^2) \int_0^k \int_0^x (x-y) dy dx = \frac{k}{3}.$$

Thus, from (16),

$$C = 1, \quad \dots (23)$$

as is to be expected for this distribution.

(b) *Normal distribution.* Standardizing without loss of generality, the frequency function

$$dF(x) = (2\pi)^{-1} \exp(-\frac{1}{2}x^2) dx \quad (-\infty \leq x \leq \infty)$$

has mean difference, from (15),

$$\Delta = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^x \exp\{-\frac{1}{2}(x^2 + y^2)\} (x-y) dy dx.$$

Making the polar transformation

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$

we obtain

$$\begin{aligned} \Delta &= \frac{1}{\pi} \int_0^{\infty} r^2 \exp(-\frac{1}{2}r^2) dr \int_{-\pi/4}^{\pi/4} (\cos \theta - \sin \theta) d\theta \\ &= \frac{1}{\pi} \cdot \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \cdot 2^{\frac{3}{2}} \\ &= 2/\pi^{\frac{1}{2}}. \end{aligned}$$

Thus, from (16),

$$C = (3/\pi)^{\frac{1}{2}} \doteq .98, \quad \dots (24)$$

a result which has previously been given by Burt (1953), who obtained it much earlier. A slightly modified form of C_n for the normal distribution has been obtained by Terry (1952) in finding the correlation between two statistics (see section VI, example 3).

(c) *Gamma distribution.* A gamma distribution with parameter $m(>0)$ has frequency function

$$dF = \frac{1}{\Gamma(m)} e^{-x} x^{m-1} dx; \quad (0 \leq x \leq \infty)$$

and (15) gives

$$\Delta\{\Gamma(m)\}^2/2 = \int_0^{\infty} \int_0^x e^{-(x+y)} x^{m-1} y^{m-1} (x-y) dy dx \quad \dots (25)$$

Making the polar transformation

$$\begin{aligned} x &= r \cos^2 \theta \\ y &= r \sin^2 \theta, \end{aligned}$$

we obtain, for the right-hand side of (25),

$$\begin{aligned} & \frac{1}{2^{2m-1}} \cdot \int_0^\infty e^{-r} r^{2m} dr \cdot \int_0^{\pi/2} (\sin 2\theta)^{2m-1} \cos 2\theta d(2\theta) \\ &= \frac{1}{2^{2m-1}} \cdot \Gamma(2m+1) \cdot \frac{1}{2m} = \frac{\Gamma(2m+1)}{m2^{2m}}. \end{aligned}$$

Thus, since $\Gamma(m+1) = m\Gamma(m)$, (25) gives

$$\Delta = \frac{m\Gamma(2m+1)}{2^{2m-1}\{\Gamma(m+1)\}^2}. \quad \dots (26)$$

Using the reduplication formula for gamma functions,

$$\pi^{\frac{1}{2}} \Gamma(2m+1) = \Gamma(m+\frac{1}{2}) \Gamma(m+1) 2^{2m},$$

we obtain from (26)

$$\Delta = 2m \cdot \Gamma(m+\frac{1}{2}) / \{\pi^{\frac{1}{2}} \Gamma(m+1)\}.$$

and finally, using (16) and remembering that $\sigma = m^{\frac{1}{2}}$,

$$C = (3m/\pi)^{\frac{1}{2}} \Gamma(m+\frac{1}{2}) / \Gamma(m+1). \quad \dots (27)$$

If we use Stirling's approximation for the gamma functions in (27) we find that

$$\lim_{m \rightarrow \infty} C = \left(\frac{3}{\pi}\right)^{\frac{1}{2}} \doteq .98, \quad \dots (28)$$

agreeing with (24), as it must, since the gamma function tends to normality with increasing m . On the other hand, as m tends to zero, (27) shows that

$$C \sim (3m)^{\frac{1}{2}}$$

so that

$$\lim_{m \rightarrow 0} C = 0. \quad \dots (29)$$

Thus, given any ϵ , we can make $C < \epsilon$ by choosing m small enough. Any gamma frequency function with $m < 1$ is unbounded as $x \rightarrow 0$, so that the necessary condition for C to be zero, stated in section III above, is here fulfilled. At the same time, for any fixed m in $0 < m < 1$, the gamma distribution shows that $f(x)$ may be unbounded with C quite large: we shall see below, e.g., that for $m = \frac{1}{2}$, $C = .78$.

It remains only to examine the rapidity with which C increases from zero as m increases. Using (27) we obtain the values:

m	$\frac{1}{2}$	1	2	3	4
C	$\frac{6^{\frac{1}{2}}}{\pi} = .78$	$\frac{3^{\frac{1}{2}}}{2} = .87$	$\frac{3^{\frac{3}{2}}}{4 \cdot 2^{\frac{1}{2}}} = .90$	$\frac{15}{16} = .94$	$\frac{35 \cdot 3^{\frac{1}{2}}}{64} = .95$

Thus the value of C increases rather quickly towards its limiting value (28). The case $m = \frac{1}{2}$ is the lowest value of general statistical interest, corresponding to a chi-squared distribution with one degree of freedom.

V. THE SAMPLE CORRELATION BETWEEN VARIATE-VALUES AND RANKS

The correlation coefficient C_n is a *parameter* of a bivariate distribution. Suppose now that we calculate the corresponding *statistic* in a single sample, i.e., we calculate the correlation r between variate-values and ranks in the sample. Clearly, r is a consistent estimator of C : for large n , it converges in probability to the value C . Thus the asymptotic results of sections III and IV hold for r as well as for C . The exact result (10) for C_n does not carry over to r , which is a ratio of random variables whose exact distribution is not easily obtained.

VI. THE ASYMPTOTIC RELATIVE EFFICIENCIES OF TESTS AGAINST NORMAL ALTERNATIVES

(24) may be used to establish, in a unified manner, the asymptotic relative efficiencies against normal alternatives of several distribution-free test statistics which are essentially efficient estimators with variate-values replaced by ranks. For this purpose we recall several theorems.

A well-known theorem due to Fisher (1925) states that the correlation C between an efficient estimator t_1 and any other estimator t_2 is equal to the square root of the estimating efficiency of t_2 . Recently, Stuart (1954) has shown that a measure of the asymptotic relative efficiency of two consistent tests, T_1 and T_2 , proposed by Pitman (1948), is equivalent to the use of the estimating efficiencies on the null hypothesis of those transformations of T_1 and T_2 , t_1 and t_2 respectively, which are consistent estimators of the parameter whose value characterizes the null hypothesis. It follows from Fisher's theorem that if t_1 is efficient, the asymptotic relative efficiency of T_2 , compared to T_1 , is C^2 .

Daniels (1944) defined a class of correlation coefficients which includes the product-moment correlation and the two rank correlation coefficients. For any coefficient r_1 of this class, Daniels showed that if the variates are separately transformed in any arbitrary fashion, the correlation on the null hypothesis between r_1 and its equivalent after transformation r_2 (and hence between any constant multiples of them) is equal to the product of the correlations between the original and transformed values for the two variates separately.

Example 1: Rank correlation coefficients as tests against trend in normal regression. Consider the product-moment correlation r_1 for a sample of n , and transform one set of variate-values into ranks, giving a new coefficient r_2 . Such a procedure is appropriate when testing against trend in one variate, the second being regarded as fixed. If the second variate is spaced at equal intervals, r_2 is in fact a constant multiple of Spearman's rank correlation coefficient r_s , and in virtue of section V the correlation between r_1 and r_s is, by Daniels's theorem, asymptotically equal to the correlation C between the variate-values and ranks. If a multiple of r_1 is an asymptotically efficient estimator of the trend parameter β , the asymptotic relative efficiency of r_s will, by Fisher's theorem, be C^2 .

For the normal regression model, the sample regression coefficient is an efficient estimator of β , so that the asymptotic relative efficiency of Spearman's coefficient is $C^2 = 3/\pi$ by (24). Since, as Daniels (1944) proved, the two rank correlation coefficients are perfectly correlated in the limit on the null hypothesis, the same result holds for Kendall's coefficient. These results were obtained directly by Stuart (1954).

Example 2: Rank correlation coefficients as tests of independence in a bivariate normal distribution. In the previous example, let both variates be replaced by ranks. The transformed coefficient is Spearman's coefficient r_s , as ordinarily used for testing independence. By Daniels's theorem, the correlation between r_1 and r_s is now C^2 , and for a bivariate normal distribution, r_1 is an efficient estimator of ρ , the correlation parameter. By Fisher's theorem, it follows that Spearman's coefficient has an asymptotic relative efficiency of $C^4 = 9/\pi^2$, and once again its perfect correlation with Kendall's coefficient gives the latter the same result. These results have been established by Hotelling and Pabst (1936) and by Moran (1951).

Example 3: Wilcoxon's two-sample test for location in normal and uniform distributions. If two normal distributions differ only by a location parameter θ , an efficient estimator of θ is $\hat{\theta} = x_1 - x_2$, on which Student's t -test is based. If variate-values are replaced by ranks, $\hat{\theta}$ becomes a linear function of Wilcoxon's (1945) two-sample test statistic, say W . The correlation between $\hat{\theta}$ and W is C , and hence, by Fisher's theorem, the asymptotic relative efficiency of the Wilcoxon test is $C^2 = 3/\pi$, as given by Pitman (1948). This is also essentially the result obtained by Terry (1952), who considered the correlation between W and a function of the $x_{(i)}$ which is asymptotically equivalent to the use of Student's t -test.

Finally, we may note that if, in this example, the distributions are uniform instead of normal, $\hat{\theta}$ is no longer an efficient estimator (Fisher (1922)) but the correlation between $\hat{\theta}$ and W is asymptotically 1, by (23), so that they are asymptotically equivalent tests. Thus the asymptotic relative efficiency of W compared to the Student t -test is 1 in this case as shown by Pitman (1948).

REFERENCES

1. Burt, C., and Morris, R. E. (1953). 'The relative merits of ranks and paired comparisons.' *Brit. J. Stat. Psychol.* VI, 112.
2. Daniels, H. E. (1944). 'The relation between measures of correlation in the universe of sample permutations.' *Biometrika*, XXXIII, 129.
3. Fisher, R. A. (1922). 'The mathematical foundations of theoretical statistics.' *Phil. Trans. Royal Soc. London, A*, CCXXII, 309.
4. Fisher, R. A. (1925). 'Theory of statistical estimation.' *Proc. Camb. Phil. Soc.*, XXII, 700.
5. Hotelling, Harold, and Pabst, Margaret Richards (1936). 'Rank correlation and tests of significance involving no assumptions of normality.' *Ann. Math. Statist.*, VII, 29.
6. Moran, P. A. P. (1951). 'Partial and multiple rank correlation.' *Biometrika*, XXXVIII, 26.
7. Nair, U. S. (1936). 'The standard error of Gini's mean difference.' *Biometrika*, XXVIII, 428.
8. Pitman, E. J. G. (1948). *Lecture Notes on Non-parametric Inference* (University of North Carolina, mimeographed).
9. Stuart, Alan (1954). 'The asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives.' *Journ. Amer. Statist. Assoc.*, XLIX.
10. Terry, Milton E. (1952). 'Some rank order tests which are most powerful against specific parametric alternatives.' *Ann. Math. Statist.*, XXIII, 346.
11. Wilcoxon, F. (1945). 'Individual comparisons by ranking methods.' *Biometrics*, I, 80.

THE DISTRIBUTION OF THE DIFFERENCE IN TOTAL RANK VALUE FOR TWO PARTICULAR OBJECTS IN m RANKINGS OF n OBJECTS

By J. W. WHITFIELD

Psychological Laboratory, University College, London

I. *Problem.* II. *Principle of Distribution.* III. *Moments.* IV. *Example.*

I. PROBLEM

A recent paper (Whitfield, 1953) discussed the problem of the total rank value assigned to one object of a series by a number of observers. It was concerned more particularly with the experimental conditions prevailing in the psychological analysis of opinion in small groups, when one individual is instructed to play a predetermined role, and every other member of the group then ranks his fellow members (including the role player) for specific characteristics. Stuart (1954) has since shown more rigorously that the assumption of a tendency to normality is justified.

For certain reasons, such as economy of experimental time or interest in the interaction of different roles, it may be desirable to have more than one predetermined role. Jonckheere (1954) discusses the general case of k roles with varying numbers of members in each. The following account is restricted to the case of two roles, with one member in each. The result agrees with the general case; and, by virtue of the numerical simplicities imposed, may prove easier for computation. Although the problem in social psychology provides a convenient example, the method is applicable in many other fields of preference or judgment.

II. PRINCIPLE OF DISTRIBUTION

Let there be m rankings of n objects; and let us suppose that the experimenter is concerned with the difference between the rankings assigned to the i th and the j th objects. In the null case i and j can be assigned any pair of rank values, in each ranking. The distribution of $r.v.i - r.v.j$ ($r.v.i$ and $r.v.j$ referring to the rank values of i and j respectively) for any one of the m rankings may then be set out as follows:

Expected Frequencies of the Various Values of $r.v.i - r.v.j$ with a Single Ranking:

Value	$(n-1), (n-2), (n-3), \dots, 2, 1, -1, -2, \dots, -(n-1)$
-------	---

Frequency	1, 2, 3, $\dots, (n-2), (n-1), (n-1), (n-2), \dots, 1$
-----------	--

Total frequency $n(n-1)$. Expected mean value, $\Sigma(r.v.i - r.v.j) = 0$.

It will be noted that, for one ranking, $r.v.i - r.v.j$ can never actually have the value 0, as no two objects can be assigned the same rank value, unless tied ranks are allowed. If tied ranks are permitted, the following argument remains substantially the same, and the test proposed remains valid. The presence of tied ranks reduces the variance of the distribution, and hence any tests of significance based on the untied case slightly over-estimates the probability of the observed value occurring by chance if ties are present.

For m rankings, the distribution is that of the sum of the $r.v.i - r.v.j$ values for each ranking, taking account of sign. This value (d) is the value $\Sigma r.v.i - \Sigma r.v.j$, and for values of m greater than 1, the value 0 can be attained. It is of interest to note, however, that for low values of m the distribution is bimodal, with modes at $+1$ and -1 . This feature rapidly disappears with increase in value of m .

III. MOMENTS

The distribution is symmetrical about zero, and hence all odd order moments are zero. Let $d =$ the sum of $r.v._i - r.v._j$, and Σd^2 correspond to the sum of the squares of all possible d values, weighted according to their probability.

It follows from the distribution that (for $m = 1$)

$$\begin{aligned}\Sigma d^2 &= 2[(n-1)1^2 + (n-2)2^2 + (n-3)3^2 + \dots (n-(n-1))(n-1)^2] \\ &= 2n[1^2 + 2^2 + 3^2 + \dots (n-1)^2] - 2[1^3 + 2^3 + 3^3 + \dots (n-1)^3] \\ &= \frac{2n}{6}(n-1)n(2n-1) - \frac{2(n-1)^2 n^2}{4} \\ &= \frac{n^2(n-1)}{6}[4n-2-3n+3] = \frac{n^2(n-1)(n+1)}{6}.\end{aligned}$$

Hence

$$\mu_2 = \frac{n^2(n-1)(n+1)}{6} \cdot \frac{1}{n(n-1)} = \frac{n(n+1)}{6}.$$

And for m rankings, $\mu_2 = \frac{mn(n+1)}{6}$.

Fourth moment (for $m = 1$)

$$\begin{aligned}\Sigma d_4 &= 2[(n-1)1^4 + (n-2)2^4 + (n-3)3^4 + \dots (n-(n-1))(n-1)^4] \\ &= 2n[1^4 + 2^4 + 3^4 + \dots (n-1)^4] - 2[1^5 + 2^5 + 3^5 + \dots (n-1)^5] \\ &= 2n(n-1)n(2n-1)\left[\frac{(n-2)(n+1)}{10} + \frac{1}{6}\right] - 2(n-1)^2 n^2 \left[\frac{(n-2)(n+1)}{6} + \frac{1}{4}\right] \\ &= 2n^2(n-1) \left\{ \frac{(2n-1)(n-2)(n+1)}{10} + \frac{(2n-1)}{6} - \frac{(n-1)(n-2)(n+1)}{6} - \frac{(n-1)}{4} \right\} \\ &= \frac{n^2(n-1)}{30} \{ 2(n-2)(n+1)(6n-3-5n+5) + 20n-10-15n+15 \} \\ &= \frac{n^2(n-1)(n+1)}{30} [2(n-2)(n+2) + 5] \\ &= \frac{n^2(n-1)(n+1)}{30} [2n^2 - 8 + 5] = \frac{n^2(n-1)(n+1)(2n^2-3)}{30}\end{aligned}$$

Hence

$$\mu_4 = \frac{n(n+1)(2n^2-3)}{30}, \text{ and for } m \text{ rankings}$$

$$\mu_4 = \mu_4 + \frac{\mu_2}{m-1} \cdot \frac{\mu_2}{1} + \frac{\mu_4}{1} \text{ where } \mu_4 = \text{fourth moment in } m \text{ ranking case,}$$

$$\mu_4 = \text{,, ,, ,, } (m-1) \text{ ,, ,, etc.}$$

$$= \mu_4 + \frac{(m-1)}{6} n^2(n+1)^2 + \frac{n(n+1)(2n^2-3)}{30}.$$

But

$$\mu_4 = \mu_4 + \frac{(m-2)}{6} n^2(n+1)^2 + \frac{n(n+1)(2n^2-3)}{30}.$$

Hence

$$\begin{aligned}\mu_4 &= \frac{mn(n+1)(2n^2-3)}{30} + \frac{n^2(n+1)^2}{6} [1 + 2 + \dots (m-1)] \\ &= \frac{mn(n+1)(2n^2-3)}{30} + \frac{n^2(n+1)^2}{6} m \frac{(m-1)}{2} \\ &= mn(n+1) \left[\frac{2n^2-3}{3} + \frac{n(n+1)(m-1)}{12} \right].\end{aligned}$$

Thus

$$\begin{aligned}\beta_2 &= \frac{36mn(n+1)}{m^2n^2(n+1)^2} \left[\frac{2n^2-3}{30} + \frac{n(n+1)(m-1)}{12} \right] \\ &= \frac{36(2n^2-3)}{30mn(n+1)} + \frac{3(m-1)}{m} \\ &= 3 - \frac{3(n^2+5n+6)}{5mn(n+1)} \\ &= 3 - \frac{3}{5m} - \frac{12}{5mn} - \frac{6}{5mn(n+1)}.\end{aligned}$$

Hence β_2 tends to 3 with increasing m .

Thus we have a symmetrical distribution with its β_2 coefficient approaching the value 3, and, therefore, a normal approximation is possible. As d is a discontinuous variable, a correction must be applied, hence

$$\text{Normal deviate} = \frac{d - \frac{1}{2}}{\sqrt{\text{var } d}} = \sqrt{\frac{d - \frac{1}{2}}{\frac{mn(n+1)}{6}}}.$$

IV. EXAMPLE

Exact tables of the distribution of d are given in the appendix for low values of m and n . Even at moderately low values of m and n , the normal approximation holds moderately well, but tends to over-estimate the chance probability. For example, an experimental group of six individuals is set up. Individuals A and B play predetermined roles; and it is predicted that the remaining members will be more favourably disposed towards A than B . The rank values given at the end of the experiment are as follows:

Ranker	Individuals Ranked					
	A	B	C	D	E	F
C	2	4	—	1	3	5
D	2	5	1	—	3	4
E	1	3	5	4	—	2
F	2	4	3	5	1	—
Total	7	16	[9	10	7	11]

$d = 16 - 7 = +9$ (in predicted direction);

$m = 4$; $n = 5$;

Variance $= \frac{4 \cdot 5 \cdot 6}{6} = 20$;

Normal deviate $= \frac{9 - \frac{1}{2}}{\sqrt{20}} = 1 \cdot 9007$;

P (for tables of the normal distribution, single tail) $= \cdot 0287$. From the exact distribution, $P = \cdot 02747$. Thus it will be seen that the approximation is reasonably exact, even with these moderate values.

It will be remembered that, when no prior hypothesis is formulated as to the direction of the difference, the probability values given in the tables will have to be doubled; if the normal approximation is employed, the two-tail tabulations must be consulted.

REFERENCES

1. Kendall, M. G. (1948). *Rank Correlation Methods*. London: Griffin & Co.
2. Whitfield, J. W. (1953). 'The distribution of total rank value for one particular object in m rankings of n objects.' *Brit. J. Stat. Psychol.*, VI, 35-40.
3. Stuart, A. (1954). 'Limit distributions for total rank values.' *Brit. J. Stat. Psychol.*, VII, 31-35.
4. Jonckheere, A. R. (1954). 'A test of significance for the relations between m rankings.' (In press.)

Distribution of the Difference in Total Rank Value for Two Particular Objects

TABLES OF EXACT VALUES OF P FOR DIFFERENCES IN TOTAL RANK VALUE

The probability values given below are for deviations in the predicted direction only, and should be doubled if absolute probability values are required.

$m = 2$

d	3	4	n 5	6	7	8
0	.63899	.59722	.57500	.56111	.55159	.54464
1	.36111	.40278	.42500	.43889	.44841	.45536
2	.25000	.29167	.32500	.35000	.36905	.38393
3	.13889	.18750	.23000	.26444	.29195	.31409
4	.02778	.10417	.15000	.18889	.22166	.24904
5		.03472	.08750	.12667	.16100	.19101
6		.00694	.03750	.07778	.11111	.14126
7			.01250	.03889	.07143	.10013
8			.00250	.01667	.03968	.06696
9				.00556	.01984	.04018
10				.00111	.00850	.02232
11					.00283	.01116
12					.00057	.00478
13						.00159
14						.00032

$m = 3$

0	.55556	.55208	.54538	.53956	.53486	.53109
1	.44444	.44792	.45463	.46044	.46514	.46891
2	.27778	.32639	.35638	.37744	.39314	.40529
3	.15278	.22222	.26750	.30011	.32483	.34421
4	.08796	.13889	.19050	.23037	.26166	.28675
5	.03241	.08333	.12788	.16993	.20492	.23393
6	.00463	.04167	.08213	.12004	.15561	.18654
7		.01620	.04763	.08167	.11432	.14513
8		.00405	.02438	.05200	.08144	.11001
9		.00058	.01050	.03067	.05537	.08131
10			.00350	.01644	.03571	.05804
11			.00088	.00778	.02166	.03986
12			.00013	.00311	.01219	.02621
13				.00104	.00624	.01638
14				.00026	.00283	.00963
15				.00004	.00113	.00526
16					.00038	.00263
17					.00009	.00120
18					.00001	.00048
19						.00016
20						.00004
21						.00000

TABLES OF EXACT VALUE OF P FOR DIFFERENCES IN TOTAL RANK VALUE

The probability values given below are for deviations in the predicted direction only, and should be doubled if absolute probability values are required.

$m = 4$						$m = 5$				
d	3	4	5	6	7	d	3	4	5	
0	·57639	·55469	·54381	·53681	·53182	0	·55658	·54662	·53857	
1	·42361	·44531	·45619	·46319	·46818	1	·44342	·54338	·46143	
2	·30633	·34578	·37269	·39170	·40580	2	·32253	·36065	·38515	
3	·20139	·25434	·29406	·32332	·34551	3	·21836	·27533	·31295	
4	·10880	·17641	·22356	·26009	·28868	4	·14120	·20072	·24695	
5	·05633	·11241	·16312	·20352	·23640	5	·07948	·13980	·18882	
6	·02546	·06727	·11312	·15453	·18950	6	·03935	·09156	·13969	
7	·00694	·03699	·07487	·11331	·14845	7	·01813	·05615	·09949	
8	·00077	·01808	·04687	·08029	·11335	8	·00656	·03202	·06801	
9		·00719	·02747	·05473	·08435	9	·00141	·01691	·04446	
10		·00217	·01482	·03571	·06100	10	·00013	·00801	·02766	
11		·00043	·00712	·02216	·04276	11		·00326	·01629	
12		·00005	·00297	·01296	·02895	12		·00107	·00899	
13			·00103	·00703	·01886	13		·00027	·00458	
14			·00028	·00349	·01175	14		·00004	·00217	
15			·00006	·00156	·00694	15		·0 ^{·4}	·00087	
16			·0 ^{·6}	·00061	·00386	16			·00031	
17				·00020	·00200	17			·00009	
18				·00006	·00096	18			·00002	
19				·00001	·00041	19			·0 ^{·3}	
20				·0 ^{·1}	·00016	20			·0 ^{·3}	
21					·00005					
22					·00001					
23					·0 ^{·3}					
24					·0 ^{·3}					

$m = 6$				$m = 7$		$m = 8$	
d	3	n	4	d	$n = 3$	$n = 3$	
0	·55766		·54362	0	·55101		·54910
1	·44234		·45638	1	·44899		·45090
2	·33689		·37166	2	·34757		·35668
3	·24173		·29239	3	·25582		·27006
4	·15959		·22161	4	·17945		·19424
5	·09960		·16104	5	·11734		·13313
6	·05716		·11213	6	·07158		·08618
7	·02879		·07441	7	·04090		·05208
8	·01310		·04678	8	·02113		·02948
9	·00525		·02763	9	·00973		·01544
10	·00156		·01526	10	·00403		·00729
11	·00028		·00780	11	·00140		·00309
12	·00002		·00363	12	·00035		·00116
13			·00149	13	·00005		·00035
14			·00052	14	·0 ^{·4}		·00008
15			·00014	15			·00001
16			·00003	16			·0 ^{·6}
17			·0 ^{·4}	17			
18			·0 ^{·3}	18			

NOTES AND CORRESPONDENCE

LIMIT DISTRIBUTIONS FOR TOTAL RANK VALUES

By ALAN STUART

Division of Research Techniques, London School of Economics

1. Whitfield (1953) has calculated the low moments of the sampling distribution of the total rank value received by one particular object when n objects are ranked by m observers, and tabulated the exact distribution for small values of m and n . For larger values of m and n , a normal approximation is given, this being justified by the fact that the excess of kurtosis tends to zero for large m .

The total rank value is, on the null hypothesis, the sum of m independent components, each of which may take the integer values 1 to n with equal probability. Stuart (1950) has shown that each such component has cumulants given by

$$\begin{cases} \kappa_1 &= \frac{1}{2}(n+1), \\ \kappa_{2r} &= B_{2r}(n^{2r}-1)/(2r), \\ \kappa_{2r+1} &= 0, \end{cases} \quad (r > 1)$$

where the B_{2r} are the Bernoulli numbers. Using the additive properties of cumulants, we have, for the total rank value,

$$\begin{aligned} \kappa_1 &= \frac{1}{2}m(n+1), \\ \kappa_2 &= mB_2(n^2-1)/2 = \frac{1}{12}m(n^2-1), \\ \text{and } \frac{\kappa_{2r}}{(\kappa_2)^r} &= \frac{2^{r-1}}{m^{r-1}r} \cdot \frac{B_{2r}}{B_2^r} \cdot \frac{(n^{2r}-1)}{(n^2-1)^r}. \end{aligned}$$

Thus, for any $n > 1$

$$\lim_{m \rightarrow \infty} \frac{\kappa_{2r}}{(\kappa_2)^r} = 0. \quad (r > 1)$$

And, since the only distribution having all standardized cumulants higher than the second equal to zero is the normal distribution (Kendall (1943), I, p. 67), we have established that, on the null hypothesis, the total rank value is asymptotically normally distributed, for any $n > 1$, with mean and variance as given.

2. In a further paper, Whitfield (1954) similarly considers the difference between the total rank values received by two objects when n objects are ranked by m observers, and again finds that the kurtosis coefficient tends to the normal value. The normal limit distribution does not follow immediately from the preceding section, since, for any finite n , the total rank values are not independent, but the result may be established as follows.¹

Consider first the difference in ranks allotted by one observer. The generating function for the probabilities of the differences is given by

$$\frac{1}{n(n-1)} \left\{ \left(\sum_{r=1}^n x^r \right) \left(\sum_{s=1}^n x^s \right) - n \right\}, \quad (1)$$

where the coefficient of x^d is the number of ways a difference of d can arise (the corrective term, $-n$, eliminates the possibility of equal ranks). If in (1) we substitute $x = e^{\theta}$, where $\theta = it$, we get the characteristic function

$$\varphi_1(t) = \frac{1}{n(n-1)} \left\{ \left(\sum_{r=1}^n e^{\theta r} \right) \left(\sum_{s=1}^n e^{-\theta s} \right) - n \right\}.$$

Summing the geometric progressions, we obtain

$$\varphi_1(t) = \frac{1}{n(n-1)} \left\{ \frac{1 - e^{n\theta}}{1 - e^{\theta}} \cdot \frac{1 - e^{-n\theta}}{1 - e^{-\theta}} - n \right\}$$

¹ The following results have also been obtained in a slightly different way by Mr. A. R. Jonckheere, of University College, London, who drew attention to the generating function (1).

$$= \frac{1}{n(n-1)} \left\{ \frac{\sinh^2(\frac{1}{2}n\theta)}{\sinh^2(\frac{1}{2}\theta)} - n \right\}.$$

The distribution of the difference between the *total* rank values is the sum of m independent components having this characteristic function.

Multiplying together the m characteristic functions, and standardizing, we obtain

$$\varphi_m\left(\frac{t}{\sigma}\right) = \left[\frac{1}{n(n-1)} \left\{ \frac{\sinh^2(\frac{1}{2}n\theta/\sigma)}{\sinh^2(\frac{1}{2}\theta/\sigma)} - n \right\} \right]^m. \quad (2)$$

We now use the Taylor series

$$\sinh x = \sum_{r=1}^{\infty} \frac{x^{2r-1}}{(2r-1)!},$$

and find that

$$\frac{\sinh^2(\frac{1}{2}n\theta/\sigma)}{\sinh^2(\frac{1}{2}\theta/\sigma)} = n^2 \frac{\left\{ 1 + \frac{1}{24} \frac{n^2\theta^2}{\sigma^2} + O\left(\frac{1}{\sigma^4}\right) \right\}^2}{\left\{ 1 + \frac{1}{24} \frac{\theta^2}{\sigma^2} + O\left(\frac{1}{\sigma^4}\right) \right\}^2}.$$

Using binomial expansions, this is, to order $\left(\frac{1}{\sigma^2}\right)$,

$$n^2 \left\{ 1 + \frac{1}{12} \frac{\theta^2}{\sigma^2} (n^2 - 1) + \dots \right\}.$$

Inserting this in the characteristic function (2), the latter becomes

$$\begin{aligned} \varphi_m\left(\frac{t}{\sigma}\right) &= \left[\frac{1}{n(n-1)} \left\{ n^2 \left(1 + \frac{1}{12} \frac{\theta^2}{\sigma^2} (n^2 - 1) + \dots \right) - n \right\} \right]^m \\ &= \left[1 + \frac{1}{12} \frac{\theta^2}{\sigma^2} n(n+1) + \dots \right]^m. \end{aligned}$$

Putting $\sigma^2 = \frac{1}{6} mn(n+1)$, as given by Whitfield (1954), we get

$$\varphi_m\left(\frac{t}{\sigma}\right) = \left[1 + \frac{\theta^2}{2m} + \dots \right]^m.$$

Thus :

$$\lim_{m \rightarrow \infty} \varphi_m\left(\frac{t}{\sigma}\right) = e^{\frac{1}{2}\theta^2},$$

which is the characteristic function of a standard normal distribution, as was to be shown.

REFERENCES

1. Kendall, M. G. (1943). *The Advanced Theory of Statistics*. London : Griffin.
2. Stuart, A. (1950). 'The cumulants of the first n natural numbers.' *Biometrika*, XXXVII, 446.
3. Whitfield, J. W. (1953). 'The distribution of the total rank value for one particular object in m rankings of n objects.' *Brit. J. Stat. Psychol.*, VI, 35-40.
4. Whitfield, J. W. (1954). 'The distribution of the difference in total rank value for two particular objects in m rankings of n objects.' *Brit. J. Stat. Psychol.*, VII, 45-49.

THE IDENTIFICATION OF FACTORS FROM DIFFERENT EXPERIMENTS

By J. A. BARLOW and C. BURT

General Problem. In a recent note in this *Journal* (VI, p. 119) Mr. Leyden has raised afresh the problem of deciding whether or not a factor obtained in one research may be legitimately identified with a factor obtained in a second research, and has briefly compared the commoner criteria put forward by different writers for this purpose. As he implies, the arguments and the evidence offered are seldom convincing; and he urges the need for "a further examination of the question."

The first systematic discussion of this problem, so far as we can discover, would appear to be that contained in a section on the 'Stability of Factors' in a paper by one of the present writers (2); but the working methods then proposed have commonly been criticized as somewhat lengthy. In that investigation, after repeating a factorization of school subjects, originally carried out in 1917, it seemed desirable to determine whether the factors reached in the later experiments were virtually the same as those found in the earlier; this raised the more general issue, namely, how far could the methods of factor analysis then adopted be relied upon to yield factors which would remain invariant, in spite of changes in persons, in tests, or in both. The principle proposed was that two factors can be regarded as identical (i) if, when the persons are the same, the two sets of factor-measurements are identical, or (ii) if, when the tests are the same, the two sets of direction cosines are identical. For algebraic proofs of the corollaries following from these assumptions, the reader was referred to earlier memoranda (cf. also this *Journal*, III, p. 124). But it may be useful to bring together into a single survey the various proofs and formulæ, and then compare the results which they yield with those obtained by the methods proposed by other writers.

Particular Forms. Three rather different cases arise.

1. First, let us suppose that we have applied the *same tests* to the *same persons* on two successive occasions, and have factorized the results. In its formal nature the problem is here practically the same as a much more familiar question—namely, that of determining how far a given factor may safely be identified with an independent or 'external' criterion. Indeed, the internal factor obtained in the first research really becomes an external criterion when we come to compare it with factors obtained from later researches. In the earliest investigation in which a summational analysis was used to verify the hypothesis of a general cognitive factor (1), one of the main objects was to show that the factor so established was virtually identical with what was commonly known as 'intelligence.' Accordingly, a direct set of ratings for examinees' intelligence was obtained from the schoolmasters; and these independent values for the 'criterion' were then correlated with what would today be called the factor-measurements. Now, when the criterion is based, not on direct assessment, but on the prior factorization of a similar set of test-measurements, it is obvious that the same procedure will still be appropriate. This means that the best measure of identity will be the correlation between the two sets of factor-measurements.

In principle, therefore, the method suggested was to apply the usual regression equation ($p = w'M$) and so calculate a factor-measurement for each of the persons tested, first from the earlier results and secondly from the later; the resulting product-sum could then be taken as indicating the amount of similarity between the two factors. Since factor-measurements are usually expressed as deviations in unitary standard measure, this product-sum will be an ordinary correlation coefficient. It may in fact be regarded as a measure of the 'reliability' (i.e., self-consistency) for one and the same factor, as assessed in two successive trials.

In practice it will be unnecessary to calculate or standardize the factor-measurements for each examinee explicitly. We have (in the usual notation)

$$p_1 p'_2 = \frac{w'_1 M_1 M'_2 w_2}{\sqrt{w'_1 M_1 M'_1 w_1} \sqrt{w'_2 M_2 M'_2 w_2}} \quad \dots (i)$$

$$\frac{w'_1 R_{12} w_2}{\sqrt{w'_1 R_{11} w_1} \sqrt{w'_2 R_{22} w_2}} \quad \dots (ii)$$

where p_1 and p_2 denote the two sets of factor-measurements and R_{12} the matrix of correlations obtained by multiplying the first set of test-measurements (M_1) by the second (M_2). In this equation the expression under either radical denotes the root-square-deviation ($\sigma_p \sqrt{n}$) of the factor-measurements, which will be 1.00 when the measurements have been standardized: the denominators are inserted to enable the computer to employ proportional weights (w_1 and w_2) if he wishes, in order to shorten the computations.

When Pearson's method of principal axes has been used, then $w'_i = v_i^{-1} f'_i$. Hence eqn. (ii) may be written

$$p_1 p'_2 = v_1^{-1} f'_1 R_{12} f_2 v_2^{-1}, \quad \dots (iii)$$

where $f'_1 f_1$ and $f'_2 f_2$ denote the sum of the squares of the factor-saturations. Burt has suggested that a fairly close approximation to this value can be obtained without calculating R_{12} , if we take

$$s_{12} = \frac{f'_1 f_2}{\sqrt{f'_1 f_1} \sqrt{f'_2 f_2}}; \quad \dots \text{ (iv)}$$

and this expression, which he calls a 'coefficient of similarity,' is simply the 'unadjusted' correlation between the two sets of factor-saturations.

It may be noted that eqn. (iv) is really a special case of the so-called 'coefficient of proportionality,'

$$R_{xy} = \frac{\sum_i r_{xi} r_{yi}}{\sqrt{\sum_i r_{xi}^2} \sqrt{\sum_i r_{yi}^2}}, \quad \dots \text{ (v)}$$

which was originally proposed as the correct form of the 'intercolumnar criterion' (a criterion at one time used by Spearman as a 'test of hierarchical arrangement': Spearman himself, however used the *adjusted* correlation). "When this 'unadjusted correlation' = 1.00, then the two columns of correlations are strictly proportional" (this *Journal*, III, p. 124; cf. *Factors of the Mind*, p. 343).

In the ideal case, in which the two sets of factor-saturations are exactly proportional, and consequently $f_1 = k f_2$ (k a scalar), we should have

$$s_{12} = \frac{k f'_2 f_2}{\sqrt{k^2 f'_2 f_2} \sqrt{f'_2 f_2}} = 1.00 \quad \dots \text{ (vi)}$$

Now, with weighted summation the direction cosines are merely the normalized values of the saturation coefficients: in the ideal case, therefore, the two sets of direction cosines will, with this procedure, also be proportional to each other. When simple summation is substituted for weighted summation, and reduced self-correlations are inserted instead of communalities of 1.00, the relation between saturations and direction cosines is no longer so direct. Nevertheless, the coefficient s_{12} , as defined by eqn. (iv), will still provide an approximate estimate of the degree to which the two sets of direction cosines coincide.

When the method of weighted summation has been used, it is possible to derive a more general criterion which will enable us to assess the resemblance of all the factors at once. We then have

$$M = FP = LV^1P, \quad \dots \text{ (vii)}$$

where L denotes the entire matrix of direction cosines obtained by normalizing the columns of F . Now in general, if we multiply two different correlation matrices of the same order, R_1 and R_2 say, the resulting product, R_3 say, will not be symmetrical. If, however, all the factors are the same, then

$$R_1 = M_1 M_1' = LV_1^1 P P' V_1^1 L' = LV_1 L', \quad \dots \text{ (viii)}$$

and similarly for R_2 . Hence

$$R_3 = R_1 R_2 = LV_1 L' \cdot LV_2 L' = LV_3 L' = R_2 R_1, \quad \dots \text{ (ix)}$$

where $V_3 = V_1 V_2 = V_2 V_1$, a diagonal matrix. When the factors are the same, therefore, the product matrix R_3 will be symmetrical. With two empirical sets of correlation tables the amount of symmetry can be estimated by correlating the totals of the columns of R_3 with the totals of the rows. This yields the so-called 'symmetry criterion' for estimating the degree of similarity between two factor-patterns (Burt, 1939; for an alternative proposal for testing pattern similarity, cf. Cattell, 1949).

2. Next let us suppose that, still keeping to the same sample of persons, we use a *different battery of tests*. On the first occasion perhaps we may have used the N.I.I.P. non-verbal tests of intelligence, and on the second occasion the Moray House verbal group tests: with each of these batteries we have (let us say) obtained a significant general factor; and we therefore want to know how far the two factors are similar. Or again, instead of changing to an entirely new set of tests, we may merely decide to drop two or three of the less effective, and try two or three different tests in their place. It is a common complaint that all such alterations in the battery are apt to change the unrotated centroid or summational factors. On the other hand, it has always been the contention of one of the present writers that the matrix of basic-plus-group factors will remain approximately invariant.

When the battery has been changed, we can no longer apply the symmetry criterion to test the similarity of the complete factor-patterns. But the procedure described above for testing the separate factors (eqns. (ii) and (iv)) can still be used. Indeed, in our view it is still the most appropriate. If we have already been able to estimate the reliability of the factors for the first trial and then for the second, we can compare the correlation between the corresponding factors as obtained on successive trials with the geometric mean of the two reliabilities: if the correlation does not differ significantly from the mean reliability, we may then legitimately infer that the two factors appear to be the same. Where there is no means of estimating their reliability separately, we shall have to take the correlation between the two sets of factor-measurements as a kind of absolute measure of the degree of similarity.

3. Thirdly, instead of using a different battery of tests, we may use a *different sample of persons*. The investigator, for example, may have administered the same set of intelligence-tests first to a group of normal children and then to a group of mental defectives; and he may then desire to ascertain

how far the factors found are the same for the two groups. Or again in another investigation he may have obtained a series of emotional assessments for groups of normal and neurotic children respectively, and he perhaps wants to know whether the general factors are identical. In these cases we can no longer compare the factor-measurements for the persons, since the two batches of persons are not the same. We are therefore compelled to compare the factor-saturations for tests.

Here the only practicable definition of identity will be that which is expressed in terms of the direction cosines; and to test their resemblance the simplest method will be to calculate the 'coefficient of similarity,' as described above (eqn. (iv)). Indeed, it was to meet the needs of this particular case that the coefficient was originally proposed. More recent writers, however, have for the most part supposed that the appropriate coefficient will be the ordinary correlation: as Mr. Leyden points out, this procedure was suggested by Fiske (*Amer. Psychol.*, III, 1948, pp. 360f.); it is the first of the devices enumerated by Cattell (*Factor Analysis*, 1952, p. 306); and still more recently it has been adopted by Eysenck in his latest book (6).

Discrepancies between Different Methods. To illustrate the difference between the two procedures we may perhaps briefly glance at the case which Dr. Eysenck himself examines, since the formula he applies has led him to draw inferences which are exactly the opposite of those previously derived by what we should hold to be the more appropriate formula.¹ In an early paper ('The unstable child,' *Child Study*, X, 1917, pp. 61-78) one of us had argued that the general factor of emotionality appeared to be the constitutional basis on which (as a result of later stress or conflict) psychoneurotic disorders most frequently developed. Many later writers, however, maintained that the general factor found among psychoneurotic children was a distinct factor of 'neuroticism.' Dr. Eysenck has accordingly been led to re-examine the evidence for identifying the two. He takes "two tables of correlations published by Burt dealing respectively with 328 neurotic or unstable children and 483 normal children" (p. 47). In the paper from which he quotes, the main conclusion was (as he rightly says), that "with the psychoneurotic group, the first two factors appear to be much the same as those obtained from normal children, except that the effects of selection are revealed by alterations in the factor-saturations and factor variances" (this *Journal*, I, p. 203). Dr. Eysenck admits that "this argument possesses a certain superficial plausibility"; nevertheless, he says, "it is difficult to see how Burt can identify the two (general) factors with each other when the correlation between them is not significantly different from zero." In any case, he adds, "the failure of the factor-saturations from the two samples to correlate significantly with each other" cannot be accounted for by "the effects of selection" (though he offers no proof for this rather dogmatic assertion). It therefore follows that both the factors cannot be identified with 'general emotionality' (as was claimed in the original paper), and possibly neither can.²

Dr. Eysenck does not give any figure for the correlation to which he refers. The ordinary 'adjusted' correlation, however, proves to be -0.33, while the 'similarity coefficient' is 0.89. It is scarcely surprising that two such different results obtained by two such different procedures should lead their respective advocates to quite opposite conclusions. In the tables which Dr. Eysenck quotes, the explanation is obvious. With both groups all the saturations for the general factor are positive and fully significant for every trait. Hence, when we calculate the two sets of factor-measurements, the minor differences in the size of the saturations (differences which in many cases are barely significant statistically) cannot greatly affect the weighted sums. However, when we compute the

¹ *Editorial Note.*—The discrepancy between these findings has already been the subject of several letters addressed to this *Journal*, chiefly from research-workers who ask for guidance in dealing with analogous problems of their own; it is hoped that this note may be accepted as an answer to their questions.

² *Loc. cit.*, p. 47. In a footnote Dr. Eysenck observes that "one's faith in Burt's identification of his factors must be considerably shaken by his failure to take account of such obvious discrepancies." But, as a matter of fact, Burt himself expressly draws attention to "obvious discrepancies" in the figures obtained, and remarks that the discrepancies "appear far more instructive than the resemblances." On an earlier page of the article, it was clearly stated that the appropriate measure of agreement in such cases is, not the ordinary correlation, but the 'unadjusted correlation'; and to make this doubly plain, the formula was explicitly given.

Dr. Eysenck seems to have supposed that, when weighting coefficients differ widely in their value, the resulting factors must also differ widely. If, however, he calculates the factor-measurements from a set of intelligence-tests, and uses a set of weights proportional to (say) .9, .8, .7, .6, . . . , and then a set proportional to . . . , .6, .7, .8, .9, then—although the weights have a perfect negative correlation—the factor-measurements will be nearly the same (applying the formula given in this *Journal*, III, p. 111, they will correlate to the extent of about + 0.8). The fact that differences in weighting, unless they are exceptionally wide, make little difference to the relative values of a series of weighted averages is emphasized in nearly every discussion on weighting procedures; and one is tempted to retort that "one's faith in Eysenck's discussion of factors must be considerably shaken by his failure to take account of such well-known facts."

ordinary product-moment correlation, we implicitly compare, not the saturations, but their deviations: we thus change the smaller saturations into *negative* deviations. And, if the incidence of these negative deviations differs in the two experiments, they may easily produce a negative correlation; they may do so, even when the measurements for the two factors are almost exactly the same.

Illustrative Example. To demonstrate the fallacy involved in substituting the ordinary for the unadjusted correlation, perhaps the most convincing method will be to take a simple but concrete example where some independent check is available. Let us start with measurements for visible objects instead of with assessments for invisible mental traits. In an attempt to obtain more familiar data to illustrate the results of a factor analysis, Thurstone and Thomson have measured boxes; and, since most boxes admit of only three linear measurements, they have had recourse to the device of combining these three-dimensional measurements in various ways, both linear and non-linear, and have then attempted to get back to the original linear measurements (e.g., Thomson, *Factorial Analysis of Human Ability*, 1948, p. 281). But such an attempt is scarcely parallel to factorial analysis as applied to human beings.

As it happens, the special interest of one of the present writers has centred on the analysis of measurements for body-build; and, for the purpose of demonstrating the method to students, he has found that the familiar medicine bottle, with stopper to represent head, provides a very plausible model. He began (A) with the first 50 medicine bottles that he found on the hospital shelves; then, to obtain specimens representing a different 'racial type,' took (B) a set of 50 bottles procured from domestic sources. Table I specifies the characteristics actually measured, and gives the saturations for the first and second factors as obtained by weighted summation in either case. With both samples the first factor appears to be a general factor for *size*, and the second a bipolar factor for *shape*.

But now let us enquire how far homologous factors from the two different samples can be reasonably identified. Consider the two general factors first. For these the unadjusted correlation between the saturations obtained from sets A and B respectively turns out to be 0.988. If, however, we adopt Dr. Eysenck's procedure, and calculate the ordinary adjusted coefficient, the correlation turns out to be negative, namely, -0.197 (cf. Table II). He would therefore affirm that "it is difficult to see how the two factors can be identified"; and would conclude that both the factors cannot be identified with size and possibly neither can.

In this experiment, however, we have a readily measurable external criterion which should settle the matter. The initial data gave vertical and cross-sectional measurements for the body of the bottle and the neck. But by size we must presumably understand tri-dimensional magnitude. The easiest way to measure size (defined in this way) is to find the volume of water each bottle will hold. The direct assessments so obtained can then be correlated with the calculated measurements deduced for the first factor. The correlations of the two sets of factor-measurements with this external criterion prove to be 0.99 for set A and 0.97 for set B. In both cases, therefore, we are clearly justified in our identification of the general factor with size.

TABLE I. FACTOR ANALYSIS OF MEASUREMENTS OF GLASS BOTTLES

Measurement	Set A		Set B	
	I	II	I	II
Height	.81	.19	.49	-.46
Width	.73	.22	.61	.34
Thickness	.65	.12	.58	.29
Length of neck	.67	-.24	.47	-.39
Diameter of neck	.61	-.38	.54	.06
Contribution to variance (per cent.)	48.6	6.0	29.2	11.5

TABLE II. ALTERNATIVE METHODS OF DETERMINING SIMILARITY OF FACTORS

Formula	AI & BI	AII & BII	AI & AII	BI & BII
Standardized product-sum	.988	.223	.000	.000
Correlation coefficient	-.197	.219	.731	.967

Incidentally it may be noted that if, following Dr. Eysenck's procedure, we calculate the adjusted correlation between factors I in set B and II in set A, we find that the coefficient is definitely positive (over + 0.4). His method would thus imply that the *general* factor in the second set resembles the *bipolar* factor in the first set much more closely than it resembles the general factor in the first set—a conclusion which does not seem even "superficially plausible."

Now let us turn to the second pair of factors. On correlating the two bipolar factors obtained from sets A and B respectively, we find a 'similarity coefficient' of only 0.223; and, since we are now dealing with positive *and* negative saturations whose sum adds up to approximately zero, the adjusted correlation will be practically the same. (With simple summation the coefficients would be identical, because the saturations for bipolar factors then add up exactly to zero, and can thus be treated as deviations about their own mean: it is possibly this fact that has misled Dr. Eysenck.) The low value implies that the two bipolar factors differ appreciably. Admittedly both describe variations in shape; but the kind of shape which is typical in the two cases is by no means the same. It would appear that, next to size, the most suitable way of classifying set A (the medicine bottles) is to distinguish those which have large bodies from those which have large necks (the proportion of height to width is much the same in all); whereas with set B the most suitable classification is apparently to distinguish the tall bottles from the wide.

By way of clinching the argument, let us calculate the ordinary (or 'adjusted') correlation between the two *different* factors in the *same* sample. In set A it is 0.731, and in set B 0.967. Those who base their criterion on the ordinary correlation, therefore, would be obliged to infer that two successive factors, extracted by weighted summation from the same correlation matrix—namely, that for set B—were almost identical! If, on the other hand, we calculate the 'unadjusted' correlation, i.e., the product-sum for the two columns of saturations, then we find that, in both sets, the coefficient is exactly zero, as of course it must be if the analysis has been correctly carried out. Thus once again we see that the 'adjustment' required for calculating the ordinary correlation turns a general (all-positive) factor into a bipolar (positive-and-negative) factor, and so obliterates the most important distinction drawn by the analysis into successive factors.

Applications. This note is concerned with methods rather than results. But we may briefly summarize the more important conclusions to which they have led. (i) In general, it would seem, provided the analyses or rotations have been based on objective, not subjective, considerations, the factors reached by group factor methods show a remarkable amount of agreement, even when, owing to selection, tests and persons differ widely. (ii) The factors reached by summational or bipolar analyses show considerable agreement in their sign-patterns and therefore in the classifications they imply: they show much wider variations in the numerical results, except when the investigations have been carefully planned in advance; for the factors to agree it is not so important for the several tests to remain the same as for a proper balance of tests or traits to be preserved. (iii) The methods here described are helpful, not only in demonstrating the identity of factors, but also in studying how far the factors change as a result of external causes—e.g., mental or physical growth. An instructive example may be found in analyses of the factors underlying measurements of body-build obtained at successive years: the results indicate minor but distinctive changes in the predominance of leptomorphic, euryomorphic, and pyknomorphic tendencies during different stages of childhood and adult life (cf., for instance, Burt, C., 'Factor analysis and physical types,' *Psychometrika*, XII, 1947, pp. 171–88, Sec. on 'Constancy of factor pattern,' pp. 181f.; *id.*, with C. Banks, 'A factor analysis of body measurements for British adult males,' based on data from the Air Force. *Ann. Eugen.*, XIII, 1947, pp. 42–55; also *id.*, *Proc. Roy. Soc. Med.*, XL, 1947, pp. 307–8).

REFERENCES

1. Burt, C. (1909). 'Experimental tests of general intelligence.' *Brit. J. Psychol.*, III, 94–177.
2. Burt, C. (1939). 'The relations of educational abilities.' *Brit. J. Educ. Psychol.*, IX, 45–71.
3. Cattell, R. (1949). ' r_p and other coefficients of pattern similarity.' *Psychometrika*, XIV, pp. 279–84.
4. Burt, C. (1950). 'The influence of differential weighting.' *Brit. J. Psychol., Stat. Sect.*, III, 105–25.
5. Cattell, R. (1952). *Factor Analysis*. New York: Harper.
6. Eysenck, H. (1953). *The Structure of Human Personality*. London: Methuen.

SPEARMAN'S FORM OF CRITERION ANALYSIS

To the Editor, *British Journal of Statistical Psychology*

SIR,—In commenting on the origins of factor analysis, the reviewer of Dr. Raymond Cattell's book overlooks one very misleading statement. Dr. Cattell declares that "Spearman's demonstration in 1904 that a single factor runs through most mental tests . . . presented the first formal and adequate statement of factor analysis." Surely "the first formal and adequate statement" was Karl Pearson's "demonstration" of the method of principal axes in 1901 (*Phil. Mag.*, II, pp. 551f.). Spearman's paper presents, not the first formal statement of factor analysis, but the first anticipations of what Cattell himself calls 'criterion analysis.' Probably Cattell has derived his idea from more recent writers like Thouless and Thurstone, who seem to have read into this earlier paper the topics of later discussions, and so have ignored its really novel contributions. Thouless, quoting the 1904 experiments, says: "Spearman's method of proving the existence of a general factor was as follows. A large number of children were tested in different performances. . . . The correlations were put down in tables of which the following is one" (duly reproduced). Similarly, in explaining to French readers the development of 'Factor Analysis as a Scientific Method,' Thurstone begins by telling us that "Factor analysis was started by Charles Spearman in a paper published in 1904. . . . He arranged the inter-correlations of a set of tests in a square table. . . . At first Spearman pointed out that, when a hierarchy exists, then the columns and rows are proportional. . . . It seems strange that his early criteria for hierarchy never seemed to be considered in relation to the rank of the table" (*L'Année Psychologique*, 1952). But both these statements give quite a wrong impression of the contents of the paper.

Spearman himself tells a very different story. After describing Thorndike's investigation, which led to the view that the mind includes "a host of specialized and independent faculties," he observes: "shortly before (in 1901) there had been published a momentous investigation by (J. M.) Cattell and Wissler (*Psych. Mon.*, II)." This indeed was the first research in which correlation was applied to mental tests. Wissler's aim was "to find a means whereby the fundamental elements of *general* and *specific* ability could be isolated and valued." The distinction between 'general' and 'specific' ability, as well as the general scheme of tests, was taken from Galton; and the statistical techniques were derived from Galton's *Natural Inheritance* and Pearson's *Grammar of Science*. According to Spearman's summary, 325 individuals were tested, and 22 mental tests applied. The correlations between most of the tests were admittedly low; but the correlations between the examination marks were far higher and seemed to indicate that "whatever it is that makes for class standing holds generally for all courses."

In Spearman's view, "the chief source of the lowness of the correlations" was to be sought in the failure to eliminate "observational errors." His own experiments were modelled on those of J. M. Cattell and Wissler; but he proposed to surmount their difficulty by applying correction-formulae to the 'raw coefficients.' It is this that formed the outstanding novelty in his enquiry, and the one that aroused the "heated controversies" to which Dr. Raymond Cattell and others refer.

Since Spearman's earliest contributions are no longer easily accessible, particularly to many of your readers who live abroad, may I offer a brief summary of his leading arguments, and then propound three questions which they suggest and which (to my knowledge) have never been adequately answered?

In point of fact, two articles were published by Spearman in 1904. They were intended for German readers, but in the end were printed in the *American Journal of Psychology*. The first dealt with the 'measurement of association,' and the sequel with the 'measurement of general intelligence' (*loc. cit.*, XV, pp. 72-101, 201-93). In the first he observes that, so far, "psychologists, with scarcely an exception, never seem to have become acquainted with the brilliant work carried on by the Galton-Pearson school" (this must mean 'German psychologists,' since Cattell and Thorndike in America and Sully in England were already thoroughly familiar with it, and so were most of their co-workers). Accordingly, he proposes to attempt a systematic review of the main formulae developed by these two writers. His account of 'Galton's particular system' is based on *Natural Inheritance* and the earlier papers on 'Correlations.' The great merit of that system, he considers, is that it affords, as Galton claimed, "a means of determining and measuring the common underlying cause of the variations." Galton's suggestion was that a given series of variations might be looked upon as "compounded of *two* superposed and independent systems of dispersion." Measuring variability by the squares of the probable deviations, he writes for the total variability $q^2 = c^2 + b^2$, where c represents that due to the common element and b that due to an irrelevant element (inaccuracy, etc.) specific to the particular set. Then $r^2 = \frac{c^2}{c^2 + b^2}$, and $c^2 : b^2$ as $r^2 : 1 - r^2$. Following Galton, Spearman proposes to measure 'inaccuracy' by $1 - r^2$ instead of $\sqrt{1 - r^2}$ as Wissler had done. He then discusses in some detail "the advantages of using ranks," and particularly of the "method of rank differences," a suggestion which he rightly attributes to Binet (*La Fatigue Intellectuelle*, 1898, pp. 252-61).

He ends with the proposal for eliminating by a statistical device those inaccuracies which cannot be removed by experimental means. His procedure is based on a suggestion of Wissler, who pointed

out that "the precision of a test may be estimated by correlating the successive trials" (*op. cit.*, p. 60). Calling the various tests $p, q, \dots$, etc., Spearman assumes that m successive trials have been made with p , and n with q ; and puts forward two formulæ, which he gives without proof.

(i) To estimate the 'true' correlation (r_{pq}) he starts, very naturally, by calculating the correlation ($r_{p'q'}$) between the *averaged measurements* for p and q respectively. He then proposes to correct this value by means of the *averaged correlations* of the separate measurements for p with those for q . For this purpose he suggests

$$r_{pq} = \frac{r_{p'q'}\sqrt{mn} - r_{p'q'}}{\sqrt{mn} - 1} \quad \dots (1)$$

(ii) His second formula is only applicable when $m = n = 2$. The averaged correlation ($r_{p'q'}$) is now to be divided by the geometric mean of the inter-correlations (the coefficients of precision). Slightly changing the notation, he writes

$$r_{pq} = \frac{r_{p_1q_1} + r_{p_1q_2} + r_{p_2q_1} + r_{p_2q_2}}{\sqrt{(r_{p_1p_2} \times r_{q_1q_2})}} \quad \dots (2)$$

By way of illustration he applies his formulæ to Pearson's 'intellectual gradings' for members of the same family, and contends that Pearson himself grossly under-estimates the 'true' correlation for mental inheritance: it should be, not .50, but over .80!

In his second article (the one to which presumably Cattell alludes) Spearman begins by complaining that Wundt's followers, unlike Galton's, have so far failed to relate the theoretical measurements of the Laboratory with the practical assessments of Life. Here "France has led the van." Binet's tests come much nearer to the processes of everyday life; yet he himself feels that it is wiser to keep to laboratory tests "as befits the rigor of scientific work." Unfortunately, practical psychologists, like Galton and Binet, still speak in terms of "the discredited faculties," while academic psychologists have insisted on the 'principle of intellective unity' (a phrase borrowed from Wundt). As a result, theorists like Ward, who quite rightly insists that "in place of a congeries of faculties, we must assume a single cognitive activity" ('attention'), are led to ridicule all attempts at *la psychologie individuelle*.

Spearman accordingly proposes to take the 'cardinal function' which Binet had called 'intelligence' and Galton 'general ability,' and to study the relations between this function as assessed in Life and the simpler cognitive functions that are accurately measurable in the Laboratory. He suggests "first an enquiry into its exact relation to sensory discrimination of which we hear so much, . . . and then, by the aid of the information thus obtained, to determine and measure this intelligence in an objective manner."

Unlike Wissler, he confines himself to very small groups from two schools only, and to (at most) three tests of sensory discrimination, similar to Galton's, but applied in class. For the 'estimation of intelligence' he relies either on teachers and older children or on examination marks. The correlations between the tests are *not* "arranged in a square table," but in columns following Wissler's plan. The 'raw' coefficients for both tests and examinations are of much the same order as Wissler's. At the preparatory school they are somewhat higher, but there the age of the children ranged from $9\frac{1}{2}$ to $13\frac{1}{2}$.

To assess the relation between Intelligence and Discrimination he applies his two formulæ. (A) With the 24 village children (i) the correlation between the average rank for intelligence (p) and the average rank for the sensory tests (q) was .66. The average inter-correlation was .38. Equation (1) gives $r_{pq} = 1.04$. (ii) The reliability (*précision*) was .55 and .25 respectively. Equation (2) gives

$$r_{pq} = \frac{.38}{\sqrt{.55 \times .25}} = 1.01.$$

(B) With the 22 preparatory school pupils, using examinations as a criterion, he obtains corrected correlations of (i) 1.04 and (ii) 0.96: average, exactly 1.00. His final conclusion is that "there exists a correspondence between General Discrimination and General Intelligence which works out with great approximation to one or absoluteness." Now, if the correlation between 'general ability' and discrimination is so perfect, there can (he argues) be no room for 'specific functions' for languages, mathematics, or music, as Galton had maintained: they must be "vanishingly minute." With one group, where the sole test was pitch discrimination, he used the ranking for Music in place of a second: yet what 'the majority of writers' would regard as a special aptitude of 'musical sensitiveness' proves to be identical with general ability. This, he points out, is fully consistent with the 'theorem of intellective unity,' as maintained by recent critics of the faculty school, like Wundt and Ward. American writers, he thinks, have unwittingly mixed the effects of training, practice, or interest with those of 'native capacity': it is the former that are responsible for superstitions about special faculties. Such effects can best be detected and eliminated by calculating the correlations with Age.

There is nothing in all this that resembles factor analysis. "General intelligence" is treated, not as a "factor" in the technical sense, but as a quality *independently measured* by examination

results or by teachers' assessments : in other words, it is an external not an internal criterion. No attempt is made to "prove its existence" : that is taken for granted.

Towards the end of his paper he briefly indicates the practical corollaries for what he calls Life. Teachers, he believes, need no longer hesitate to use a child's 'place in the arithmetic class' as an index of his 'brightness,' merely because certain German writers have maintained that mathematics forms an 'independent faculty.' However, it is wise to make sure whether mathematical subjects really furnish the best guide : anyhow there can be no doubt that some subjects are more reliable than others. To find the 'true rank of the respective school activities,' he calculates their 'correlations with general intelligence' after 'eliminating observational errors,' and finds—on pooling the results from all three experiments—that their values range from .99 (for classics) and .96 for pitch discrimination down to .57 for discrimination of light and weight respectively. According to 'Galton's system,' any given correlation "depends on two very different influences"—first the extent to which the process tested is identical with the ability we want to estimate, namely intelligence, and secondly "the accuracy with which we have estimated it." Relative 'accuracy' he proposes to measure in accordance with Galton's principle. Accordingly, he compares the *square* of the corrected correlation (r^2) with its complement ($1 - r^2$). Thus the 'saturation' (in Galton's metaphor) of Galton's own test of weight discrimination is only $.44^2 = .19$; hence the ratio of error to accuracy is .81 to .19, or roughly 4 to 1. For practical purposes, he concludes, a test of discrimination obtained with "a few minutes with a monochord" is probably the best, since it is independent of age, industry, practice, and general education.

From these 'pedagogical conclusions' he turns to another topic recently raised by Galton—the possibility of finding a 'rational basis for public examinations' (particularly those newly introduced for the Government service). Spearman believes it should be possible to arrange the main types of post in a hierarchical order—from those involving the most complex functions like 'administering provinces' down to the simplest clerical or manual tasks. His validity coefficients (as we should call them) provide a method for ranking examination-subjects in the same serial fashion. And his final conclusion is that, by this procedure, "the value of any method of examination for any given post can be ascertained, since it depends on (b) the hierarchical intellectual rank of the test and (c) the hierarchical intellectual rank of the duties involved in the post."

Both the idea and the name of a 'hierarchy,' as applied in this way to intellectual activities, seem to have originated with French writers. Comte, in his *Cours de philosophie positive* (1842), had suggested the view that the main departments of knowledge could be arranged in a 'hierarchy' with mathematics at the top. Binet, on many points a follower of Comte, also writes of the *hiérarchie des intelligences diverses*. Spearman probably got the word from Pearson, who had fully discussed Comte's conception in the last chapter of his *Grammar of Science*, and had argued that "the best form of classification would be a branching tree," not a linear series as Comte assumed. Spearman believes that a careful study of the available data would support the serial view, since, no matter which 'branch of school study' we take as a criterion, the order of the correlations is practically the same. Here again his problem is really one of validity ; and the point he seeks to prove is that, no matter what type of intellectual proficiency you wish to predict, you will always find Classics more accurate than Mathematics, and Mathematics than English subjects, and so on. Spearman thus plainly favours the Comtean conception of the *scala intellectus*, and expresses "no small surprise" at finding Classics rather than Mathematics at the top of the list.

It is here that he introduces the little table referred to by Thouless and Thurstone. He gives, not several square 'tables,' as Thouless implies, but only one. His sample of children does not comprise "a large number," but only 22. Nor are the coefficients "inter-correlations between a set of tests" ; they are inter-correlations between marks for school examinations, with the one best test included for comparison. Moreover, they are not actual correlations for marks or ranks, but correlations for the differences between mark-order and age-order (p. 250). The idea of arranging correlations in the form of a square determinant or matrix was not Spearman's, but Pearson's. The idea that they would steadily diminish as the chief common cause became more and more 'diluted' was Galton's, though Spearman may have derived it from Pearson's exposition. Finally, Spearman nowhere points out, as Thurstone declares, that "the rows and columns are proportional" : on the contrary, he supposes that the decrements are roughly the same. "Et voilà justement," as Voltaire was fond of remarking, "comme on écrit l'histoire !"

Spearman's correction-formula would in fact not hold good of a table in which "the rows and columns are proportional." His original conception of a 'perfectly uniform arrangement' was one in which the figures descend, not by equal proportions, but by equal steps : compare, for example, the 'perfect' hierarchical table printed in *Brit. J. Psychol.*, VIII, 1916, p. 175. Equation (2) is correct only on this latter assumption. The same is true (as Garnett showed) of Spearman's formula for the intercolumnar correlation : with an equi-proportional table the required criterion would be the inner product, not the product-moment correlation. The early histories of factor analysis (e.g., Brown's) make it clear that the first attempt to fit a table of observed correlations by constructing a 'hierarchy' (in the later sense of a matrix of rank one) was that described in Burt's 1909 article (*Brit. J. Psychol.*, III, pp. 161, 162). It is his tables that contemporary writers like Thorndike, Freeman,

Simpson, and Garnett reprint when they wish to illustrate the method of fitting a table of observed correlations with the expected values to be derived from a given hypothetical factor. And when Spearman mentions the 'proportionality criterion,' he calls it 'Burt's equation.' But, seeing how often he criticized the factorial approach adopted by Burt and later by Thurstone, and preferred what some might regard as the more solid ground afforded by a direct external criterion, it is surely unfair to saddle him with the very procedure he so consistently rejected.

Spearman's own account of his earlier work is given in his Appendix to *The Abilities of Man*; and there, for his 'original proof,' he refers us, not to the pages dealing with the 'square table,' but to "pp. 268-272," which are concerned with equations (1) and (2) cited above. His subsequent procedures (the equations 'for detecting specific correlation') were developed from these same correction-formulae. There is no fitting of hierarchies, but once again a reliance on independent 'reference values.' In short, his own procedure from first to last is essentially a form of 'criterion analysis,' not of 'factor analysis.' 'Criterion analysis,' so Professor Cattell explains, is a "method of statistical verification" where the correspondence between (a) a set of 'independent variables' (the tests or other predictors) and (b) a set of 'dependent variables' (together forming the external criterion we wish to predict) is determined by "including both in a common correlational matrix"; and that is precisely what Spearman attempted.

But could results so obtained possess more than a theoretical interest? The practical inference (that, because Discrimination has a corrected correlation of unity with Intelligence, therefore we may safely use tests of Discrimination to predict Intelligence) does not seem to follow. To determine practical validity it would surely be better to adopt a suggestion originally put forward by Pearson, in criticizing Spearman's formula, namely, that we should estimate the maximum hypothetical correlation between p and q by taking weighted averages of the measurements or assessments, and determine the weights by least squares, as in the theory of multiple correlation. The actual calculation would follow the method illustrated by Thomson (this *Journal*, I, p. 30). Using the same values as Spearman, I find that this yields, for the correlation between Intelligence and Discrimination, a maximum hypothetical correlation of .51.

And this brings me to the questions I wish to put. Your reviewer mentions 'criterion analysis as a "conceivable alternative": which, if any, of the foregoing formulae would he accept? And could he, or some other reader, explain the following points? (i) How were Spearman's equations derived? (ii) Why did he choose to identify intelligence with Discrimination? Binet, whom he so often quotes, had identified it with Attention (cf. *L'année psychologique*, VI, 1899, pp. 248f.), in common with most other advocates of 'intellective unity,' like Wundt and Ward. (iii) Finally, why do Cattell and the others I have mentioned suppose that Spearman's 1904 article provides "the first formal and adequate statement of factor analysis"?—Yours, etc., JACQUES LAFITTE.

Reply.—I regret that in my review the space available prevented me from dealing adequately with the notion of 'criterion analysis.' How far that is the same as Spearman's earlier method is rather a question for Dr. Cattell. But neither seems to be a method of 'analysis.'

My replies to M. Lafitte's three questions would be briefly as follows:

(i) Brown (*Mental Measurement*, 1911, p. 86) suggests that Spearman's idea of averaging and comparing inter-correlations and cross-correlations was derived from Pearson's references to correlation between "like parts" and "unlike parts." For eqn. (2) Spearman offers a 'practical verification' in the form of a 'target experiment,' borrowed from Galton's experiment (*Natural Inheritance*, p. 69). Both writers seemed to have based their formulae on semi-intuitive reasoning. Thomson, however, supposes that Spearman was describing an *actual* experiment, and comments on the remarkable coincidence that the theoretical value (1.00) should have been so exactly obtained; he himself got values ranging from .68 to 1.37 (*Mental Measurement*, p. 157). Later Spearman says that "for the new form a new proof was derived from Yule's well-known formula for partial correlation."

But for this and all such formulae the most direct proofs are reached by the method adopted in *Mental and Scholastic Tests* (p. 271). This regularly starts with linear expressions for each observed variate, putting (say) $p' = c + a + d$ and $q' = c + b + e$, where c denotes the element common to all tests, a and b the elements (if any) common only to one pair or the other, and d and e chance errors. These expansions may then be substituted in Pearson's product-moment formula. The proofs so obtained have the merit of exhibiting the assumptions requisite to justify the working formula.

(ii) In 1904 the two English text-books that dominated the teaching of psychology in this country were those of Bain and Sully; and in both Discrimination rather than Attention was taken as the characteristic cognitive activity. Bain, in discussing 'intellectual' processes, describes 'Discrimination' as the 'fundamental property'; Sully declares that "the discernment of difference is the most fundamental and constant element in all intellection; it is known as Discrimination. . . . All intellectual ability is in general determined by the perfection of the senses with the closely conjoined function known as Discrimination." Spearman repeats this statement almost verbatim in formulating his final conclusion. His reasons for rejecting Binet's view are indicated in *The Nature of Intelligence*, pp. 11f.

(iii) Recent writers who wish to summarize Spearman's earlier work generally turn to the Appendix in Spearman's book, not to the original article. In the paragraphs quoted by M. Lafitte, Spearman gives the modified version of his 'corrective equation,' not that actually used in the 1904 paper; and he re-states the main conclusion in factorial terms: "the factor common to the two sensory discriminations showed itself functionally identical with that common to the two (ratings for) intelligence." This, however, as M. Lafitte points out, was quite different from showing that the same general factor runs through all or most mental tests. But in the short historical accounts given by later writers like Thurstone and Cattell the two propositions have evidently become confused.

Thouless in the last edition of his book has modified the passage quoted above. He states that "the problem (of factor analysis) and the method of solving it had been formulated . . . by Karl Pearson," and cites the paper on 'Lines of Closest Fit.' Instead of Spearman's table of correlations between examination marks, he now gives a "table of inter-correlations between 5 tests given to 757 subjects," and a "best fitting hierarchy" calculated by the method proposed by Burt in his 1909 article. Thurstone is mistaken in supposing that "the early criteria were never considered in relation to rank." The fact that Spearman's 'tetrad difference criterion' was the ordinary determinantal criterion for a matrix of rank one was perfectly familiar to Yule (from whom Spearman took the term 'tetrad'), to Maxwell Garnett, and to W. F. Sheppard (who pointed out the theoretical possibility of using the ordinary determinantal criteria for higher ranks).—F. N. HARPER.

A RELIABILITY COEFFICIENT BASED ON ANALYSIS OF VARIANCE

To the Editor, *British Journal of Statistical Psychology*

SIR,—I am endeavouring to standardize a battery of eight intelligence-tests for children of school age, and wish first to estimate its reliability. A recent paper refers to "a factorial method of assessing reliability by means of analysis of variance, developed for use by the International Institute Examinations Inquiry," and describes it as "a device that is specially suitable when it is impracticable to apply the tests twice either in the same or in a parallel form." What is the formula?

—Yours, etc., A. F. MAHMOUD

Reply.—The formula proposed may be expressed most simply by the equation

$$r_{tt'} = 1 - \frac{1}{F},$$

where $F = \frac{\text{mean square between groups}}{\text{mean square within groups}}$, the so called 'variance ratio' tabulated by Snedecor. As F approaches infinity, $r_{tt'}$ approaches 1; when F is 1, $r_{tt'} = 0$. For working purposes it is simplest to take

$$r_{tt'} = \frac{\text{mean square within groups} - \text{mean square between groups}}{\text{mean square within groups}} = \frac{F - 1}{F}.$$

The quantities required for this ratio can be obtained most easily from the formula

$$r_{tt'} = \frac{n\bar{r}_{ij}}{(n-1)\bar{r}_{ij} + 1} = \frac{\sum_{i,j} \sum_{i,j} r_{ij} (r_{ii} = \bar{r}_{ij})}{\sum_{i,j} \sum_{i,j} r_{ij} (r_{ii} = 1)}.$$

The novice can calculate the requisite ratio by using the 'pooling square,' putting the self-correlations all equal to 1.00 in the diagonal subsquares (for the denominator), and all equal to the average inter-correlation ($\bar{r}_{ij}$) in the non-diagonal subsquares (for the numerator).

In my memorandum, appended to the Report of the Committee, set up by the I.I.E.I., on *Marks of Examiners* (1936, pp. 245–312), it was proposed to assess reliability by the proportional contribution of the 'general factor' to the total variance, i.e., $r_{tt'} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$; and this can be most readily estimated by the method described above. For the relations between the formulæ derived by analysis of variance, factor analysis, and by correlating tests in the usual way, see *Factors of the Mind*, p. 275. For a worked example see 'The reliability of teachers' assessments,' *Brit. J. Educ. Psychol.*, XV, pp. 80–92. As there pointed out, the same formula (or its algebraic equivalent) can be used when the constituent 'tests' are questions or other items marked either right or wrong (pass or fail), and is then equivalent to the average of the split half reliability coefficients based on all possible splits.

—CYRIL BURT.

BOOK REVIEWS

The Study of Behavior: *Q*-Technique and its Methodology. By William Stephenson. Chicago: University Press. 1953. Pp. x + 376. £2 16s. 6d.

The aim of Dr. Stephenson's book is to describe what he believes to be a 'new methodology.' In his opening chapter on 'Prolegomena to *Q*,' he explains that the hypothetico-deductive method, as employed by Hull, Eysenck, and others, is to his mind inadequate: their own rejection of all other procedures he compares to the Irishwoman's claim that her son alone was in step and everybody else out. Equally he is dissatisfied with Burt's treatment of induction: it "leans on nineteenth century rather than twentieth century thinking" (why he regards Keynes, Broad and Fisher as nineteenth century thinkers is not made clear). His own view is that "the inductive principle must be reduced to a conclusion about a factor based on *prior* expectancies": it must not be "used as an absolutist principle which, having given rise to factors in a simple structure, leaves the factorist to explain them *a posteriori*, by hindsight" (a sound nineteenth century doctrine: see Jevons, *Principles of Science*, 1877, chap. XII).

His exposition of the new methodology opens with a detailed discussion of the differences between '*Q*-technique' and 'Burt's *P*-technique.' Among those psychologists who have adopted factorial methods "none," he says, "with the notable exception of Burt, has given any attention to the methodological implications. . . . Ward alone reached in a penetrating manner into *R*, and was highly critical in consequence." "*Q*-technique, we guessed at once, was the necessary and sufficient schema that Ward looked for and could not find." The reference seems to be to Ward's footnote (*Psychological Principles*, pp. 433-4) criticizing Stern's *Differentielle Psychologie* (1911). Now Stern had drawn a clear distinction between correlating traits and correlating persons—between *inter-individuelle* and *intra-individuelle Korrelation*, as he called it, and expounded (with the aid of diagrams like those of Cattell) the various kinds of problems to which each is more appropriate. Strange to say, Stephenson nowhere mentions Stern's discussion, though it provided a useful terminology adopted by later writers, including Stephenson himself.

Burt, he says, like Spearman, Thurstone, Holzinger, and the rest, is still preoccupied with "inter-individual differences": *Q*-technique is concerned with "intra-individual significance." Burt only correlates persons when the factors could just as well be obtained by correlating traits. As evidence, he quotes a long passage from what he says is "a statement by Burt." But, as a matter of fact, the statement was not written by Burt at all: it comes (as the reference shows) from Babington-Smith's attempt to "make Stephenson's position clearer." If Stephenson will refer to Margaret Davies's historical review of previous work on 'Correlating Persons' (*Brit. J. Psychol.*, XXIX, 1939, pp. 404-21), he will find numerous early analyses which are certainly 'intra-individual,' some even relating to changes within a single person only. Take, for example, the experiments on artistic appreciation carried out in collaboration with Margaret Bulley, Heather Dewar, Violet Pelling, and others, in which one or more persons ranked 50 pictures: does Dr. Stephenson suppose that these could have been factorized just as well by *R*-technique?

However, there are two further differences which he now brings forward. First, he says, a distinctive feature of the new method is the '*Q*-sort': the subject is given, say, 50 cards, and is told to sort them into nine piles (to be allotted 1, 2, . . . , 9 marks) containing (say) 2, 4, 6, 8, 10, 8, 6, 4, 2 cards in each. Yet is this really such a novel device? Secondly, *Q*-technique alone can deal with problems of 'dependency.' He reprints Professor Kendall's 'family tree' representing 'dependency analysis' and 'interdependency analysis' as two distinct offspring of 'multivariate analysis.' Kendall puts 'factor analysis' as one of the descendants of the latter. That, says Dr. Stephenson, is true of *R*-technique and *P*-technique, which is merely an '*R*-variate design.' But *Q*-technique belongs to the other branch: it is the child of a marriage between 'dependency factor analysis' and 'Fisherian methodology.' It thus imparts to the investigation far more freedom and far more objectivity.

He next turns to the differences between his procedure and the various alternatives described by Raymond Cattell. As Dr. Cronbach has elsewhere observed, "Stephenson's cheerful belligerence is at least matched by the gusto with which R. B. Cattell has entered the arena"; and Dr. Stephenson here replies: "Cattell," he says, "has proliferated into *P*, *T*, *O* and other techniques." But Cattell's version of them is merely 'actuarial.' Like Burt's, on which they are based, they "merely adumbrate variate designs in *R*-technique: *R* and *Q* alone have the status of methodologies." Cattell is preoccupied with 'external frames of reference'; whereas what the study of behaviour requires is an 'internal frame of reference.' And accordingly Dr. Stephenson shows "how one of Cattell's tables should have been analysed, viz., from the internal standpoint of the persons who made the assessments rather than from that of the persons about whom the assessments were externally made."

The second part of the book deals with 'practical applications.' It is not so much a review of results already achieved as an illustrated programme of experiments which might prove fruitful. Here we are told how 'Q-methodology' should throw light on clinical, social, type- and self-psychology, and how it may be applied to data obtained with projective tests or questionnaires. These chapters indeed form the most interesting portions of the whole volume.

Throughout his book Dr. Stephenson is full of new and original ideas. He is not always clear or convincing; but he has a lively and entertaining style. And those interested in novel possibilities for factorial work will find his many suggestions well worth study and reflection. *Charlotte Banks*

The Psychology of Personality. By Bernard Notcutt. London: Methuen & Co. 1953. Pp. 260. £1 1s.

In his preface Professor Notcutt tells us that the purpose of his book is "more to evaluate than describe the activities of psychologists"; and he therefore has attempted to choose those current "methods and concepts" that "have the best promise for the future."

In the discussion of methods, what will chiefly interest the statistical reader will be the chapters on factor analysis and the logic of validation. In the latter Professor Notcutt very aptly observes that "so far, the subjective psychologists have had most of the interesting ideas but have been unable to prove them: the objective psychologists have made delicate instruments of analysis, but applied them to trivial problems." He himself believes that, "by using tests of specific prediction, we can hope to combine the advantages of both subjective and objective methods."

Factor analysis he contrasts with typology. Typology, we are told, follows 'a zoological analogy' and tries to sort human beings into separate pigeon holes; factor analysis follows a geometrical analogy and attempts to interpret the inter-correlations as the expression of 'factors of the mind.' But in geometry you can choose any axes you like and there is an infinite number to choose from. Consequently, "during the thirties in America when factor analysis seemed a new device, new factors were two a penny on every bramble"; one factorist spent most of his time denying the very existence of a hypothetical ability that some rival factorist had picked as the cornerstone of his theory. However, by now, he thinks, "the more acute controversies are dying down. . . . A majority of workers on both sides of the Atlantic seem prepared to accept Burt's four factor theory: a general factor, group factors, specific factors, and residual factors of error. . . . These four correspond to the Aristotelian description of genus, species, proprium, and accidens." Factors in short are now seen to be, not 'real entities' planted like phrenological organs in the brain, but 'principles of classification.'

In his discussion of concepts, Professor Notcutt is less convincing. His chapters on 'basic motives,' 'psychodynamics,' 'patterns of maturation,' and the like, are too exclusively based on the modernized, or perhaps one should say the Americanized, version of Freud's psychoanalytic doctrines. Nevertheless Professor Notcutt has a special relish for what he calls 'the debunking type of research'; and merrily makes game of the psychiatrist's attempt to "classify normal individuals in terms of the diseases they have so far managed to escape."

Older readers may perhaps be a little irritated by the grammatical lapses, the frequent misspellings (particularly of foreign names and terms), and the way in which certain theories and methods are stated in caricature to make criticism clearer and objections more amusing. The long lists of references are most useful; but, when one wants to learn more of Professor Notcutt's new method of validation, it is annoying to find that the date, the volume, and page of the periodical in question are invariably omitted. On the other hand, the youthful student, for whom after all the book is chiefly intended, will enjoy the author's gift of vivid and colloquial phrasing and his eager avoidance of anything like pedantic proof. *Barbara Wilson*

Introduction to the Theory of Games. By J. C. C. McKinsey. New York: McGraw-Hill. 1952. Pp. x + 371. 55s.

Nearly a decade has elapsed since von Neumann and Morgenstern published their book on *Theory of Games and Economic Behaviour*. In its special field it took rank almost at once as a classic. The work of Keynes and his school had dealt mainly with the principles governing *mass* action in the economic sphere, e.g., the relations of total income, general consumption, investment, and the like; von Neumann and Morgenstern endeavoured to start with the behaviour of *individuals*, and such temporary coalitions as they might be led to form.

The basic problem underlying their theory can be formulated quite simply: how, under various specified conditions, is the individual to gain a maximum satisfaction from his activity when the activity of other participants can only be guessed or tentatively inferred? This is not unlike the problem discussed by Edgeworth, by means of a far simpler mathematical model, in his tract on

'Mathematical Psychics.' The device adopted by von Neumann and Morgenstern was to regard an individual's choice at each successive stage as analogous to a move in a game. This treatment brings the questions of economic behaviour much nearer to those studied by the psychologist. But, as the writers pointed out, their results have implications in many other fields besides those of finance: they throw an unexpected light on military strategy, on political theory, and even on the social psychology of everyday life. Moreover, Wald has shown that the logic of statistical inference can be fruitfully treated as an analysis of a 'zero-sum two-person' game. The 'theory of games' is thus full of concepts, methods, and new problems which should be of special interest to the psychologist. For example, it is highly instructive to invent games for children of a particular age, and examine how far the outcome departs from that predicted by the theory, and what are the causes—cognitive, affective, or conative—for such departures. It is therefore a matter for some surprise that work of von Neumann and his followers has hitherto attracted so little notice from psychological investigators.

Dr. McKinsey's book endeavours to incorporate recent developments in the subject. It is "intended to be used as a college text-book for a course for upper division and graduate students"; and, though it is not his purpose to draw special attention to the psychological possibilities, his chapters should be both intelligible and suggestive to those students who are familiar with the elementary treatment of maxima, minima, and convergence. One of the great attractions of the theory, even at the start, was the fact that, in the main, no advanced knowledge of mathematical techniques was required. On the other hand, the reasoning employed is often lengthy and exacting. For the most part, Dr. McKinsey's exposition will be found easier to follow, though perhaps less exciting, than that adopted by the earlier writers. Quite apart, therefore, from its special aim, this is a book which should do much to stimulate an active interest in the subject, particularly among students devoting themselves to psychological research.

Cyril Burt

Notes on the Teaching of Statistics in Schools. By B. C. Brookes. London: Heinemann. 1952. Pp. 8 + 80. 5s.

Forty years ago, in his book on *The Teaching of Algebra*, Sir Percy Nunn strongly urged that "an elementary study of statistics should be regarded as part of the normal programme of secondary school mathematics"; and in the companion volume intended for the pupil, *Exercises in Algebra*, he provided practical demonstrations based on his own teaching experience; since so many of the examples are concerned with mental testing, school marks, and even the concept of 'factors,' this section still provides an admirable introduction to statistical methods for students of education and psychology. On both theoretical and practical grounds he was thus able to make a strong case for his main contention. As he pointed out, during the opening years of the century statistical procedures had, "quite suddenly, come to play a prominent part in industrial, commercial, and political life, and even begun to spread into various branches of physics and biology." To this, it might be added that the commonest forms of modern reasoning are essentially statistical; and the inductive method itself is made to rest on the theory of probability. Nevertheless, the very notion of inferences by probability still remains almost unknown to the ordinary educated man.

It is therefore scarcely a matter for surprise that a Committee appointed by the Royal Statistical Society, after examining the statistical teaching in universities and university colleges, turned its attention to the teaching of statistics in schools. Mr. Brookes's 'Notes' were prepared to assist the discussions of the Committee, and are introduced by a Foreword from Professor Egon Pearson, who was the Committee's chairman.

The first section describes how the author, himself a teacher of some experience, has succeeded in dealing with the subject at an elementary level. It contains a list of appropriate topics and excellent suggestions for the collection of suitable material, not only from school work in science, but also from the ordinary everyday activities of young people. The second section discusses the relevant syllabuses adopted by the various boards examining for the General Certificate of Education. Among educationists who have more recently advocated the introduction of statistics into the school curriculum, there has been some doubt whether it should figure as a branch of the more abstract courses in mathematics or whether it should be linked more closely to the concrete subjects, e.g., biology or civics. In the main, the practice of those examining bodies that recognize the subject is apparently to regard it as a special option under the general heading of mathematics. However, since the teaching of statistics in schools is still an experimental venture, it is clearly desirable that for the present the requirements should be left as elastic as possible. Meanwhile, stimulated very largely by the interest aroused during their training as teachers in applications of statistical methods to educational and psychological problems, an increasing number of teachers are now attempting to pass on their enthusiasm to the pupils in their charge; and there can be little question that their enterprise deserves all the encouragement and assistance it can receive. Mr. Brookes's 'Notes' form an admirable contribution towards this end.

Cyril Burt